

# **Inference Generation and Reading Disability**

**Leesa Jayne Clarke**

**PhD**

**Psychology**

**September 2009**

## **Abstract**

This thesis investigated on-line inference generation in 7 to 10-year-old children. Using the Simple View of Reading (Gough & Tunmer, 1986) as a framework to classify reading disability, poor comprehenders and poor decoders were compared with chronological age-matched typical readers.

In Experiments 1, 2, 3 and 4 a self-paced reading paradigm was employed to assess on-line inference generation. Reading times to target sentences that were consistent and inconsistent with inferred context were compared. Experiments 1 and 2 showed that children made inferences about fictional characters' emotion-states and spatial inferences on-line, indexed by increased reading times to inconsistent target sentences. Results indicated that poor comprehenders made fewer inferences on-line than either typical readers or poor decoders.

In Experiment 3, participants generated inferences based on early-acquired general knowledge on-line but not inferences based on later-acquired knowledge. This was replicated in Experiment 4 for early-acquired knowledge, however only poor comprehenders failed to generate inferences on-line using later-acquired knowledge. There was evidence of carry-over in post-target sentences, which was less pronounced in poor comprehenders.

In Experiment 5, a self-paced listening paradigm showed that poor comprehenders generated fewer on-line inferences than typical readers did when processing spoken language, but contrary to prediction so did poor decoders.

In Experiment 6, children read short vignettes followed by a question. They then evaluated four answers varying in plausibility. Poor comprehenders were equally accurate but responded faster than either poor decoders or typical readers. This was consistent with the view that poor comprehenders apply a lower standard for coherence when processing language.

Together, the results support the view that decoding and linguistic comprehension dissociate, as predicted by the Simple View of Reading. These results contribute further evidence, using an on-line measure, that poor comprehenders generate fewer inferences than typical readers or poor decoders, and suggest this may be driven by the application of a lower standard for coherence when comprehending written or spoken language.

## **Table of Contents**

<b>Abstract</b>	<b>i</b>
<b>Table of Contents</b>	<b>ii</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>x</b>
<b>Acknowledgments</b>	<b>xi</b>
<b>CHAPTER 1</b>	<b>1</b>
<b>Reading and Comprehension</b>	
1.1 What is Successful Reading Comprehension?	1
1.2 Components of Successful Reading Comprehension	3
1.2.1 Basic Skills	3
1.2.1.1 Decoding and Word Recognition Skills	3
1.2.1.2 Listening Comprehension	5
1.2.1.3 Vocabulary	5
1.2.1.4 Suppression	7
1.2.1.5 Working Memory	7
1.2.2 Higher Level Comprehension Skills	11
1.2.2.1 Understanding Story Structure	11
1.2.2.2 Comprehension Monitoring	12
1.2.2.3 Inference Generation	13
1.2.3 Moderating Influences on the Process of Comprehension	15
1.2.3.1 Exposure to Print	15
1.2.3.2 Domain Knowledge	16
1.2.3.3 Other Moderating Influences	16
1.2.4 Summary	17
1.3 Reading and Comprehension Difficulties: The Simple View	17
1.4 Children with Poor Comprehension Skills	20
1.4.1 General Cognitive Ability in Poor Comprehenders	21
1.4.2 Vocabulary in Poor Comprehenders	22
1.4.3 Working Memory in Poor Comprehenders	24
1.4.4 Language Skills in Poor Comprehenders	25

1.4.5	Poor Comprehenders' Understanding of Story Structure	29
1.4.6	Comprehension Monitoring in Poor Comprehenders	29
1.4.7	Identification of a core deficit in Poor Comprehenders	31
1.5	Inference Generation and Reading Comprehension	33
1.5.1	Classification of Inference Types	34
1.5.2	Measurement of Inference Generation	35
1.5.2.1	Off-line Measures of Inference Generation	35
1.5.2.2	On-line Measures of Inference Generation	36
1.5.3	The Time Course of Inference Generation	37
1.5.4	Summary	39
1.6	Children's Inference Generation	40
1.6.1	Inference Generation in Poor Comprehenders	41
1.6.2	Summary	44
1.7	Aims of this thesis	45
 <b>CHAPTER 2</b>		<b>47</b>
<b>Reading Disability and Children's On-line Generation of Emotion-state and Spatial Inferences</b>		
2.1	Introduction	47
2.2	Experiment 1 Method	53
2.2.1	Materials	53
2.2.1.1	Experimental Stories	53
2.2.1.2	Comprehension Questions	56
2.2.1.3	Standardised Tests of Reading and Language Ability	58
2.2.2	Design	60
2.2.3	Participants	61
2.2.4	Procedure	64
2.3	Results	65
2.3.1	Comprehension Questions	65
2.3.2	Reading Times to Target Sentences	67
2.4	Discussion	70
2.5	Experiment 2	72
2.6	Method	73

2.6.1	Design	73
2.6.2	Materials	73
2.6.2.1	Experimental Stories	73
2.6.2.2	Comprehension Questions	75
2.6.3	Participants	76
2.6.4	Procedure	79
2.7	Results	79
2.7.1	Comprehension Questions	79
2.7.2	Reading Times to Target Sentences	81
2.8	Discussion	84
2.9	Summary and Conclusion	85

## **CHAPTER 3 88**

### **Knowledge Integration and On-line Inference Generation**

3.1	Introduction	88
3.2	Experiment 3 Method	91
3.2.1	Design	92
3.2.2	Materials	92
3.2.2.1	Experimental stories	92
3.2.2.2	Comprehension Questions	94
3.2.2.3	Standardised Tests of Reading and Language Ability	95
3.2.3	Participants	96
3.2.4	Procedure	99
3.3	Results	99
3.3.1	Comprehension Questions	99
3.3.2	Reading Times to Target Sentences	102
3.4	Discussion	104
3.5	Experiment 4	106
3.5.1	Introduction	106
3.6	Experiment 4 Method	107
3.6.1	Design	107
3.6.2	Materials	108
3.6.2.1	Experimental Stories	108

3.6.2.2	Comprehension Questions	110
3.6.2.3	Standardised Tests of Reading and Language Ability	110
3.6.3	Participants	112
3.6.4	Procedure	115
3.7	Results	116
3.7.1	Comprehension Questions	116
3.7.2	Reading Times to Target Sentences	118
3.7.2.1	Syllable Reading Times	118
3.7.2.2	Inhibition Scores	119
3.7.3	Reading Times to Post-target Sentences	120
3.7.3.1	Syllable Reading Times	120
3.7.3.2	Inhibition Scores	121
3.8	Discussion	122
3.9	Summary and Conclusion	125

## **CHAPTER 4** **126**

### **The On-line Generation of Emotion-State and Spatial Inferences in the Auditory Modality**

4.1	Introduction	126
4.2	Method	127
4.2.1	Design	127
4.2.2	Materials	128
4.2.2.1	Experimental stories	128
4.2.2.2	Comprehension Questions	129
4.2.2.3	Standardised Tests of Reading and Language Ability	129
4.2.3	Participants	130
4.2.4	Procedure	133
4.3	Results	134
4.3.1	Comprehension Questions	134
4.3.2	Listening Times to Target Sentences	137
4.4	Discussion	139

## **CHAPTER 5** **144**

### **Reading Disability, Passage Coherence and Sensitivity to Context**

5.1 Introduction	144
5.1.1 Standard for Coherence and Inference Generation	144
5.1.2 Comprehension Monitoring and Standard for Coherence	147
5.1.3 Understanding Story Structure and Standard for Coherence	147
5.1.4 Rationale for Experiment 6	147
5.2 Experiment 6 Method	149
5.2.1 Design	149
5.2.2 Materials	150
5.2.2.1 Item Generation	150
5.2.2.2 Experimental Scenarios	150
5.2.3 Participants	152
5.2.4 Procedure	155
5.3 Results	156
5.3.1 Yes responses	156
5.3.2 Response Times	158
5.3.3 Response Times to Uncertain Answers	159
5.4 Discussion	161

## **CHAPTER 6** **165**

### **General Discussion**

6.1 The Simple View and Experimental Findings	165
6.1.1 Overview of Findings	168
6.1.2 Theoretical Context	173
6.2 Differences Between Off-Line and On-Line Processes	174
6.2.1 Individual Differences and the Development of Comprehension Difficulties	175
6.2.1.2. Individual Differences, Inferencing Difficulties and Comprehension	175

6.2.2 Decoding Difficulties	178
6.3 Implications	179
6.4 Limitations	182
6.5 Suggestions for Further Research	184
6.6 Conclusions	186
<b>References</b>	<b>188</b>
<b>Appendices</b>	
Appendix 1 – Experiment 1 Test Materials	206
Appendix 2 – Experiment 2 Test Materials	217
Appendix 3 – Experiment 3 Test Materials	226
Appendix 4 – Experiment 4 Test Materials	236
Appendix 5 – Experiment 6 Test Materials	243



<b>List of Tables</b>	<b>Page</b>
Table 1.1 Developmental sequence of inference types in narrative comprehension	41
Table 2.1 Hatcher grading and WORD reading age equivalents for experimental stories	54
Table 2.2 Reasons for participant withdrawals from study	62
Table 2.3 Characteristics of participants by group	63
Table 2.4 Means (and Standard Deviations) of scores to responses to questions asked at the end of each story	66
Table 2.5 Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story	67
Table 2.6 Comparison of inhibition in syllable reading times by group	70
Table 2.7 Hatcher grading and WORD reading age equivalents for experimental stories	75
Table 2.8 Characteristics of participants by group	77
Table 2.9 Means (and Standard Deviations) of scores to responses to questions asked at the end of each story	80
Table 2.10 Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story	81
Table 2.11 Comparison of inhibition in syllable reading times by group	83
Table 3.1 Hatcher grading and WORD reading age equivalents for experimental stories	94
Table 3.2 Characteristics of participants by group	97
Table 3.3 Means (and Standard Deviations) of scores to responses to questions asked at the end of each story	100
Table 3.4 Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story	101
Table 3.5 Comparison of inhibition in syllable reading times by group	104
Table 3.6 Hatcher grading and WORD reading age equivalents for extended experimental stories	109
Table 3.7 Characteristics of participants by group	113

Table 3.8	Means (and Standard Deviations) of scores to responses to questions asked at the end of each story	116
Table 3.9	Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story	117
Table 3.10	Comparison of inhibition in syllable reading times by group	119
Table 3.11	Comparison of inhibition in syllable reading times by group	121
Table 4.1	Characteristics of participants by group	131
Table 4.2	Means (and Standard Deviations) of scores to responses to questions asked at the end of each story.	135
Table 4.3	Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story	136
Table 5.1	Example scenarios with question and possible responses	151
Table 5.2	Characteristics of participants by group	153

<b>List of Figures</b>	<b>Page</b>
Figure 1.1 Classification of reading disorders based upon the Simple View of Reading	19
Figure 2.1 Mean syllable reading speeds to target sentences in milliseconds	68
Figure 2.2 Mean syllable reading speeds to target sentences in milliseconds	82
Figure 3.1 Mean syllable reading speeds to target sentences in milliseconds	102
Figure 3.2 Mean syllable reading speeds to target sentences in milliseconds	118
Figure 3.3 Mean syllable-reading speeds to three post-target sentences in milliseconds	120
Figure 4.1 Mean listening times to target sentences minus time taken for the audio to play in milliseconds	138
Figure 5.1 Percentage 'yes' responses for each answer type by group	157
Figure 5.2 Mean response times in milliseconds to each answer type as a function of group	158
Figure 5.3 Reaction times of responses to uncertain answers by valence of response	160
Figure 5.4 Reaction times to responses in uncertain conditions by	160

## **Acknowledgments.**

I would like to thank a number of people that have helped me over the years spent working towards my PhD. First and foremost I would like to thank Steve, for his support, tolerance and for always believing in me when I have doubted myself. Without his patience, encouragement, and at times powers of persuasion, I would not have completed this thesis.

I would also like to thank my supervisor, Maggie Snowling, for helping, guiding and supporting me. I feel very privileged to have worked with her. I would also like to thank Claudine Crane and Andy Young for their help and guidance. I would like to thank the ESRC for choosing to fund my research.

I am indebted to all of the wonderful children who took part in this research, and the staff of the schools I have worked in who made me feel so welcome during my time there. Thanks are also due to Pamela Baylis for helping me to screen participants, and to Paula Clarke and Emma Truelove from the ReadME project. I would also like to thank Michelle St. Clair, Sophie Brigstocke and everyone at the CRL who has shown me friendship and support.

Huge thanks are also due to Lindsey Holroyd, for making coming back to work after maternity leave so much less stressful than I expected. Knowing that Malachi is happy while I'm working has made a huge difference.

Finally, I would like to dedicate this thesis to my son Malachi, who came along half way through my studies, and helped me to remember that life does not revolve around working all hours and studying for a PhD.

# CHAPTER 1

## Reading and Comprehension

### *1.1 What is Successful Reading Comprehension?*

There is much more to reading than decoding printed words. Obtaining a full and coherent understanding of a text requires readers to extract meaning from words in the context of sentences and paragraphs, to integrate it with prior knowledge of the world and to elaborate upon the explicit information given. Only when all of these things are achieved can reading be said to be truly successful.

Caccamise and Snyder (2005) define a good reader as “one who reads with deep understanding, as demonstrated by such things as the ability to abstract, apply or generalize the information in the text”. Perfetti, Landi and Oakhill (2007) observe that the ultimate goal in the acquisition of reading comprehension is the ability to understand written text as well as one understands spoken language.

According to van Dijk and Kintsch (1983), three levels of complex processes are involved in successfully comprehending a text. The first of these is described as the linguistic level, which involves word decoding and recognition, and this is an absolute prerequisite for reading comprehension (Kintsch, 1998). At the second text base level, the reader begins to unravel the meaning within the text by extracting its explicit meaning. Meaning units or propositions are merged to form the locally coherent microstructure, and this microstructure is further organised into key topics or themes forming a globally coherent macrostructure (Kintsch & Rawson, 2007). This stage may necessitate the generation of some coherence inferences, in order to resolve pronouns and anaphora (Kintsch, 1998). At the third and final level is the situation model. This is a dynamic mental representation constructed from the meaning elicited from the text as reading proceeds, plus any prior knowledge that is activated and

integrated with it at any stage in the reading process. According to Kintsch, successful reading comprehension can be equated to the successful construction of a rich and elaborate situation model (Kintsch, 1998; van Dijk and Kintsch, 1983).

Gernsbacher (1990, 1995) proposed that construction of the situation model operates via a dynamic Structure Building Framework in which a core infrastructure is constructed, and new propositions encountered are mapped onto these foundations. If they are coherent with previously integrated information, they are mapped onto the existing structure, if this is not plausible then an additional branch or substructure is added. According to this view, a mental representation is analogous to the construction of a mind-map (e.g. Buzan, 1974) with ideas linked onto a main topic, and projections added to represent the incorporation of new ideas.

If Kintsch's view of successful reading comprehension is adopted (Kintsch, 1998), complex processes at the text, sentence and word levels are implicated, involving the generation of inferences. A level of prior knowledge is required to enable the generation of elaborative inferences and also for some additional coherence inferences to be made during reading. Previous experience will determine the precise content of this knowledge base, which includes a diverse selection of information about concepts such as spatial and temporal relations, object properties, knowledge of cause and effect in relation to events or actions, and plans that motivate a person's actions or emotions (Graesser, Millis and Zwaan, 1997). It is important too that readers know how and when to perform inferential and integrative processes when reading a text (van Dijk and Kintsch, 1983). An understanding of text structure and comprehension monitoring skill is also critical at this higher-level stage in the comprehension process (Cain, Oakhill and Bryant, 2004).

From the outline above, it is clear that reading comprehension is a complex process that involves co-ordination of many skills, and impairment in any one of these has the potential to impact on successful comprehension. The development and integration of these component skills depends upon the adequate development of a range of genetically and environmentally determined cognitive abilities.

## *1.2 Components of Successful Reading Comprehension*

### *1.2.1 Basic Skills*

Basic skills and resources implicated in reading comprehension are word recognition, vocabulary, listening comprehension, suppression of irrelevant information, and working memory. These feed into higher-level skills such as comprehension monitoring, inference generation and understanding of story structure. Ultimately, integration of these component skills, each dependent on factors such as motivation, attention, domain knowledge and reading goals, will determine the level of comprehension achieved. In the sections that follow, the relationship between reading comprehension and component skills will be discussed.

#### *1.2.1.1 Decoding and Word Recognition Skills*

The ability to decode and identify words is a necessary but not sufficient prerequisite for successful reading comprehension (Perfetti, Landi & Oakhill, 2007). Gough, Hoover and Peterson (1996) describe an elegant illustration of how decoding can proceed without comprehension. They tell the story of how John Milton taught his daughters to read Greek and Latin classics to him, to compensate for his failing sight, yet despite being able to read the words, the girls did not understand Greek or Latin.

However, it is clear that comprehension will be inhibited if decoding is laboured or inaccurate (Snyder, Caccamise & Wise, 2005). In beginning readers,

word decoding is a significant impediment to reading comprehension, and children's spoken language comprehension naturally exceeds their written language comprehension. In a study by Curtis (1980) it was found that the correlation between scores obtained in tests of reading and listening comprehension was considerably lower (0.50) in 7 to 8 year old children than it was in 8-10 year old children (0.75), reflecting how decoding limits reading comprehension in early reading. By the age of 9-10 years (fourth grade), it has been reported that readers' emphasis is more focused upon extracting information from text, than on decoding the words (Keenan et al, 2006). Similarly Willson and Rupley (1997), and Rupley, Willson and Nichols (1998) observed that the contribution of decoding to variance in reading comprehension decreases with age. In their longitudinal study, Catts, Hogan and Adlof (2005) report a subgroup of 'late emerging poor readers' who show reading problems around the fourth grade, perhaps when extracting meaning from the text becomes more demanding, and children tackle more difficult texts.

Perfetti (1985) proposed that decoding bottlenecks are the proximal cause of poor reading comprehension. However, contradictory to this, there is an abundance of evidence demonstrating that decoding and comprehension skills can dissociate, and that comprehension skill can fall behind relative to word identification skill (e.g. Yuill & Oakhill, 1991; Nation & Snowling, 1997, 1998a, 1998b; Catts, Hogan & Adlof, 2005).

It could be argued, however, that results of tests of reading accuracy do not give any indication of the proportion of processing capacity allocated to word recognition (Cain & Oakhill, 2006b; Perfetti, 1985). Even though reading may be accurate, it is plausible that poor comprehenders may recognize words more slowly or effortfully, and may therefore have reduced processing capacity in reserve for



comprehension. However, this seems unlikely, given that poor comprehenders achieve equivalent scores to typical readers on tests that include a fluency measure (e.g. Cain, Oakhill & Bryant, 2000; Stothard & Hulme, 1995).

#### *1.2.1.2 Listening Comprehension*

As decoding ability improves, listening and reading comprehension become increasingly highly correlated (Curtis, 1980) until an asymptotic level is reached where listening and reading comprehension are at broadly similar levels (Gernsbacher, 1990). Support for this is provided by Adlof, Catts, Hogan and Little (2005) who, based on a longitudinal study of 604 typically developing second grade readers, showed that word recognition as measured by tests of single word reading and narrative reading, accounts for most of the variance in reading comprehension. By eighth grade, reading comprehension and listening comprehension in the same children were indistinguishable. In a study of young adult readers, Gernsbacher (1990) reported correlations of 0.90 between reading comprehension and listening comprehension.

The advancement of reading comprehension skill is constrained by a child's level of spoken language comprehension. In typical readers, once word recognition is relatively automatized, listening comprehension and reading comprehension levels are positively correlated (de Jong & van der Leij, 2002). However, children with poor word identification skills usually perform at normal levels in tests of listening comprehension, which does not predict reading comprehension in this group (Curtis, 1980).

#### *1.2.1.3 Vocabulary*

Converging evidence strongly suggests that vocabulary size accounts for a significant proportion of variance in comprehension skill (Seigneuric, Ehrlich, Oakhill

& Yuill, 2000; Rosenshine, 1980). Cunningham and Stanovich (1997) showed that level of receptive vocabulary in first grade had far reaching effects, and remained a modest but significant predictor of reading comprehension in 11<sup>th</sup> Grade. In another study, kindergarten children's receptive vocabulary was shown to predict reading comprehension in fourth and seventh grades (Snow, Tabors & Dickenson, 2001). Longitudinal work suggests that the influence of vocabulary on reading comprehension increases as the contribution made by word recognition diminishes over time in developing readers (e.g. Rupley, Willson & Nichols, 1998).

Vocabulary training has been shown to have a positive effect that generalises to reading comprehension (e.g. Beck, Perfetti and McKeown, 1982; Kameenui, Carnine & Freschi, 1982; McKeown, Beck, Omanson & Perfetti, 1983), however in some interventions, only vocabulary-specific improvements are observed (Mezynski, 1983; Pany, Jenkins & Schrek, 1982). Nagy (1988) suggests two reasons why vocabulary training may fail to boost reading comprehension; firstly, the training may not provide sufficient elaboration to facilitate new word learning in context, and secondly, the training may focus on words that are not critical to overall passage comprehension (Nagy, Anderson & Herman, 1987; Freebody & Anderson, 1983). It is plausible that when vocabulary training does impact on comprehension, the training involves work on higher-level comprehension skills such as inference generation and comprehension monitoring, suggesting that the link between vocabulary and reading comprehension may be an indirect one. For example, in a training study, Nash and Snowling (2006) found that in children with poor language skills, comprehension was boosted by a vocabulary intervention that taught the use of semantic mapping as a technique for inferring meaning from context. Similar gains in comprehension were not observed in a comparison group who were simply taught definitions of the same words.

Exposure to print is an important predictor of vocabulary expansion in developing readers (Nagy & Anderson, 1984). A child with an unremediated reading difficulty is likely to read less, with Matthew effects leading to further deterioration in vocabulary relative to peers. In line with this, Stanovich (1993) found that high exposure to print was associated with better vocabulary knowledge in children with poor comprehension skills. Interestingly, Stanovich (1993) also showed that higher comprehension skills in the presence of low exposure to print were associated with high cognitive ability, suggesting that children with higher IQ are better able to compensate for less than ideal conditions when comprehension skills are developing, and that exposure to print and the resulting increase in vocabulary size may mediate the effects of low cognitive ability on comprehension.

#### *1.2.1.4 Suppression*

According to Gernsbacher and Faust (1991), the ability to suppress the activation of irrelevant information is an important component of comprehension. During reading, activation occurs to all plausible meanings of words encountered in text, and the reader has to use evidence from context or syntax to narrow these down in order to identify the appropriate meaning. Gernsbacher and Faust found that speed of suppression of activation of inappropriate homophones was related to comprehension skill.

#### *1.2.1.5 Working Memory*

Integration and storage of incoming text are key activities performed during text comprehension. These processes are likely to place demands on working memory, however the precise nature of the relationship between memory skills and comprehension has yet to be specified. Working memory has been shown to correlate positively with reading comprehension level at a variety of ages in developing readers

(Cain, Oakhill & Bryant, 2004; Seigneuric, Ehrlich, Oakhill & Yuill, 2000; Leather & Henry, 1994).

Kintsch and Kintsch (2005) suggest that working memory is important for comprehension because it determines the number of connections a reader is able to make between concepts presented in a text. If two concepts presented cannot be held in working memory at the same time, they suggest that new links between them cannot be formed and that this limits the development of the situation model and textbase. Supportive evidence for this is provided by Yuill, Oakhill and Parkin (1989) who found that poor comprehenders were less able than controls to use resolving information to solve an anomaly in a passage if the resolving information was further away from the anomaly in the text. Similarly, if inconsistencies are further apart in a text and higher demands are therefore placed on working memory, poor comprehenders are less likely to detect them than controls (Oakhill, Hartt & Samols, 2005).

Ericsson and Kintsch (1995) propose that richer and more elaborate situation models enable the formation of retrieval structures that tap into long-term memory and facilitate integration of incoming information with the existing knowledge base. Thus, if poor working memory limits the formation of connections between premises within a text, it effectively exerts a cumulative effect on comprehension – less elaborate situation models are less formed, and this reduces the number of retrieval structures available for knowledge integration, which in turn limits comprehension of further incoming text. In line with this idea, Oakhill, Cain and Yuill (1998) suggest that working memory limitations may result in an inability to build complete situation models, and thereby limit readers to the integration of local text only. Available

working memory resources may be taken up with lower level text processing such as word identification and syntactic processing (Oakhill, Cain & Yuill, 1998).

Verbal working memory span has been shown to correlate positively with reading comprehension skill (Daneman & Carpenter, 1980) and inferencing skills (Singer, Halldorson, Lear & Andrusiak, 1992). Swanson and Berninger (1995) established that participants with good comprehension skills performed better on a battery of verbal working memory tasks including story retelling, auditory digit sequencing, sentence span and phrase sequencing. They also demonstrated that working memory differences linked to comprehension levels were specific to the verbal domain and that comprehension skills were not linked to differences in phonological short-term memory. In support of this, a number of studies report that verbal working memory is a better predictor of reading comprehension than other components of working memory (Cain, Oakhill & Bryant, 2004; Cataldo & Oakhill, 2000; Nation, Adams, Bowyer-Crane & Snowling, 1999).

Although evidence cited so far seems to indicate that working memory and comprehension ability are positively related, there are a number of studies that cast doubt on this conclusion. Some studies failed to find a relationship between comprehension skill and performance in tests of digit span, sentence recall and serial recall (Oakhill, Yuill & Parkin, 1986; Stothard & Hulme, 1992). Furthermore, word length and lexicality effects in serial recall were not found to relate to comprehension level (Nation et al, 1999).

A distinction may need to be made between working memory capacity and working memory efficiency, and it has been suggested that comprehension is only predicted by tasks that require information to be stored and processed simultaneously (Cain, Oakhill & Bryant, 2004; Cain, Oakhill & Lemmon, 2004). Furthermore, it is

also likely that the relationship between working memory and comprehension changes over time, and that this may explain some of the contradictory evidence. In a three-year longitudinal study of comprehension development in 74 French children, Seigneuric and Ehrlich (2005) found that as the importance of word recognition diminished, the contribution of working memory capacity (measured via an adapted listening span task) increased.

It has been suggested that links between comprehension skill and performance in some verbal working memory tasks can be better explained by underlying verbal and specifically semantic skills (e.g. Nation, 2005, Nation et al, 1999; Nation & Snowling, 1998a). In a study exploring predictors of reading comprehension in 180 mixed ability children aged between 9 and 11 years, Goff, Pratt and Ong (2005) showed that language variables accounted for a higher proportion of variance in reading comprehension than memory variables. Orthographic processing measured via a test of irregular word reading accounted for 36% variance in reading comprehension after age and IQ were controlled.

In support of this, Nation and Snowling (2004) recently explored the relationship between reading comprehension and language skills in a longitudinal study of 72 typical readers. A series of hierarchical regression analyses showed that semantic skills, listening comprehension and vocabulary at age 8.5 predicted unique variance in reading comprehension at age 13 after age, nonverbal ability, nonword reading and phonological skill had been controlled. They also found that these language variables accounted concurrently for a large proportion of unique variance in reading comprehension skill at age 8.5.

In conclusion, the evidence to date does not clearly specify the relationship between working memory and comprehension skill. Some involvement of working

memory is evident, however the precise mechanism by which it is involved in reading comprehension remains unclear. There is considerable debate over whether deficits in working memory task performance are artefacts of difficulties with the language manipulations involved in performing such tasks.

### *1.2.2 Higher Level Comprehension Skills*

In addition to the basic abilities outlined above, comprehension also involves applying an understanding of story structure, comprehension monitoring, and inference generation.

#### *1.2.2.1 Understanding Story Structure*

Children are quickly conditioned to basic conventions about stories, such as the understanding that stories usually have a title, an introduction, and an ending (Perfetti, Landi & Oakhill, 2007). Most children, before reaching the age of three, become familiar with the format of stories told to them by caregivers. Typically a name for the story is provided, and then storytelling begins with standard openings such as ‘once upon a time...’ and ends with something like ‘and they lived happily ever after’. On reaching school age, most children are sensitive to these and more subtle aspects of story structure. As literacy is acquired, reading experience also adds to the development of this understanding (Perfetti, 1994).

Using story-telling tasks, Cain (2003) and Cain & Oakhill (1996) showed that comprehension skills were positively related to production of coherent stories that followed typical structural conventions. Typically, less able comprehenders produced stories that contained fewer linguistic features that enabled concepts to be integrated smoothly.

Training in story structure understanding has been shown to boost reading comprehension (Yuill & Joscelyne, 1988; Gordon & Braun, 1985) and in Yuill and

Joscelyne's (1988) second experiment, improvements in comprehension following training were specific to poor comprehenders, implying that the prerequisite knowledge was already activated by typical readers.

#### *1.2.2.2 Comprehension Monitoring*

Periodically when reading text, competent readers check whether the incoming information fits with their current representation of the situation. When an inconsistency is detected, they use a range of strategies to achieve some sort of resolution of the anomaly. There may be different outcomes that the reader settles upon following a process of reanalysis. Possibilities include development of a satisfactory explanation that can be integrated with the mental representation and will account for the anomaly; acceptance of the anomaly as some sort of error in the text; identification and reparation of a previous reading or comprehension error, or decision to ignore the anomaly or put it to one side in the expectation that some explanation will follow. Reading behaviours that typically reflect effective comprehension monitoring include repetitions, slowing of reading, hesitations and self-corrections (Paris & Myers, 1981). Studies of comprehension monitoring typically employ error detection tasks, and manipulate factors such as error placement, type, and presence and/or position of resolving information. The proficiency and frequency of comprehension monitoring behaviours has been shown to vary widely both within and between individuals, and has been linked to overall reading comprehension performance (Oakhill, Hartt & Samols, 2005; Zinar, 2000; Yuill, Oakhill & Parkin, 1989) and word identification skills (Hacker, 1997; Paris & Myers, 1981). Levels of comprehension monitoring are positively related to overall comprehension skill (Cain, Oakhill & Bryant, 2004).



A number of explanations have been posited for reduced accuracy or frequency of comprehension monitoring behaviours, including working memory deficits, difficulties with word identification, lack of awareness of the need to engage in comprehension monitoring, lack of motivation to read for meaning (deSousa & Oakhill, 1996), and acceptance of a low standard of coherence. It is possible that poor comprehension monitoring shares a reciprocal relationship with inference generation; in short a reader may fail to detect inconsistencies because they do not generate inferences that enable their identification.

### *1.2.2.3 Inference Generation*

Inference generation lies at the very core of successful reading comprehension. If text were read without drawing any inferences, understanding would be extremely limited. For example take the passage:

*“Jake was in the garden. As he sat underneath the apple tree a bird flew onto the lawn. He crept up on it stealthily. After a few seconds, he pounced. The bell on his collar warned the blackbird and it flew away just in time!”*

In order for the events in this passage to make sense it is necessary to infer information not explicitly stated in the passage. To do this the reader needs to exploit prior knowledge of language conventions and the domain of interest (Kintsch & Rawson, 2007). For a rich understanding of this passage, information such as Jake was a cat, therefore he wanted to catch the bird, that the bird on the lawn was the blackbird that flew away, that Jake ran onto the lawn, and that the bell made a noise when Jake moved fast so that the bird was alerted, would need to be inferred. In terms of understanding who did what to whom, the reader would need to recognize that the 3<sup>rd</sup> person pronouns he/his refer to Jake, and that “it” refers to the blackbird. This example shows how fundamental inferences are to understanding even short and

basic texts. It also highlights the range of task demands required for the construction of different inferences.

There is considerable debate as to which inference types individuals habitually generate *on-line*, or automatically during reading. On-line inferences are considered to be made contemporaneously as reading proceeds, whereas off-line inferences are thought to be more strategic, and are constructed retrospectively after a sentence or paragraph has been read. Thus, inferences made on- and off-line differ in terms of their time course, and the extent to which they are automatic or more strategically driven.

Kintsch (1998) has suggested that the majority of inferences are generated during the construction of the situation model. If meaning cannot be extracted from the text without anaphor and pronoun resolution, then these inferences are made at the textbase level. Graesser and Zwaan (1995) propose that only a proportion of inferences are actually made during reading. Gernsbacher (1990, 1995) suggests that rather than a simple off-line versus on-line distinction; levels of activation generated during processing of incoming text may determine the off-line or on-line status of particular inference types.

There exists a range of theoretical positions from which predictions can be made about whether particular inference types are generated on-line. Two leading theoretical positions from which such predictions can be made are the minimalist view and the constructionist view. Supporters of the minimalist view proposed by McKoon and Ratcliff (1992) suggest that only inferences required for the maintenance of local coherence are made automatically on-line. Further inference generation is driven by readers' strategies or goals. This view is difficult to disprove, given that one of the tenets of this theory is that the proportion of inferences generated would vary

according to the circumstances. The constructionist position (Graesser, Singer & Trabasso, 1994) postulates that readers make all the inferences they can automatically on-line, with the aim of building a deep and enriched situation model of the text. In the constructionist view, readers strive to maintain both global and local coherence.

The relationship between inference generation and reading comprehension is complex. Overall, there is convincing evidence that poor inferential skills are linked to comprehension failure (e.g. Yuill & Oakhill, 1991). Studies have shown individual differences in the frequency of inference generation, and in whether particular inference types are generated on- or off-line, and these will be discussed later in this review.

### *1.2.3 Moderating Influences on the Process of Comprehension*

#### *1.2.3.1 Exposure to Print*

Perfetti (1994) has suggested that low print exposure may be related in a cumulative fashion to comprehension skill. Initially, low print exposure may result in impoverished understanding of story structure and convention, which in turn may affect comprehension. If comprehension is poor then reading becomes less rewarding, and therefore a child will read less widely. It is clear that some variance in word recognition skills can be accounted for by levels of exposure to print (Cunningham & Stanovich, 1998). However the evidence for a relationship between exposure to print and reading comprehension level is equivocal, particularly in the early stages of learning to read. Cain, Oakhill and Bryant (2000) compared performance of 7-8 year old poor comprehenders, typical readers matched on reading age and younger comprehension-age matched controls on an author recognition task. They found no differences between the skilled and less skilled group in their recognition of children's authors' names. The younger comprehension-age matched group had marginally

lower scores in this test, indicating that they had not had greater exposure to print than the older poor comprehenders.

Although these results seem to indicate that the poor comprehenders did not have lower print exposure, data from older children is suggestive of a relationship between exposure to print and development of reading comprehension. Stanovich, West, Cunningham and Cipielewski (1996) found that in an older sample, some variance in reading comprehension was predicted by exposure to print. They suggested an indirect relationship, such that print exposure mediates development of vocabulary, general knowledge and metalinguistic skills that impact directly on comprehension skill. Goff, Pratt and Ong (2005) provide support for this view. In their study of predictors of reading comprehension skill in 180 children aged 9-11 years, they found that print exposure measured via a children's book title recognition test made a small unique contribution to variance in reading comprehension.

#### *1.2.3.2 Domain Knowledge*

According to Caccamise and Snyder (2005), domain knowledge has a profound effect on an individual's capability to extract meaning from a text. For example, it would be difficult for someone with no prior knowledge of psychology to read and understand a journal article about models of reading comprehension. Although they may be able to decode the words in the article, it is likely that their vocabulary will not extend to knowing the meaning of specific technical terms, and they will have difficulty drawing appropriate inferences about the ideas presented. If vocabulary knowledge is adequate then the reader may be able to infer the meaning of unknown words from context, but if too many words are unknown this becomes increasingly difficult. Therefore a comprehension difficulty demonstrated during the reading of a specific text may reflect characteristics of that text rather than a general

comprehension difficulty in the reader. Ultimately, construction of a situation model cannot proceed adequately without a sufficiently elaborate appreciation of the situation!

#### *1.2.3.3 Other Moderating Influences*

There are a number of influences external to the process of comprehension that moderate levels of reading comprehension, such as attention, environmental factors such as background noise and distractions, reading goals, motivation, text type and writing style (Kintsch & Kintsch, 2005; Guthrie & Wigfield, 2005). Together, these factors will influence reading performance, to a greater or lesser extent. These will not be discussed in detail in this review; however, they must be borne in mind as potential explanations for comprehension failure.

#### *1.2.4 Summary*

It is clear that reading comprehension is a complex activity requiring the co-ordination of input from different cognitive systems. For effective comprehension, basic decoding, working memory, vocabulary and listening comprehension skills are required, and these facilitate higher level processes such as comprehension monitoring, inference generation and understanding of story structure. In turn, these processes are moderated by factors such as domain knowledge, strategy use, attention, reading goals, motivation, text type and writing style. Given all of this, it is easy to appreciate why comprehension sometimes fails. In the sections that follow, the relationship between reading difficulty and comprehension difficulty will be explored, followed by a discussion of specific comprehension difficulties in developing readers.

### *1.3 Reading and Comprehension Difficulties: The Simple View*

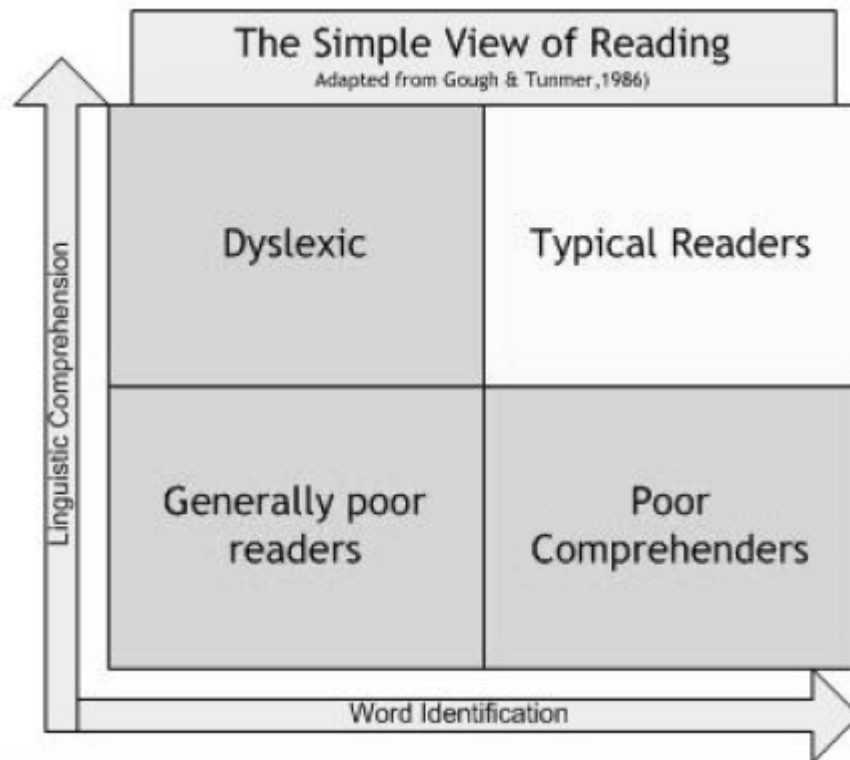
The Simple View of Reading (Hoover & Gough, 1990; Gough & Tunmer, 1986) proposes that reading comprises the product of decoding and linguistic

comprehension, each of which exerts an independent influence on reading comprehension. A multiplicative rather than additive relationship is proposed because reading cannot proceed in the absence of either decoding or linguistic comprehension. Returning to the earlier example of Milton who got his daughters to read Greek and Latin classics to him, he was unable to successfully read and extract meaning from them because his sight precluded any decoding, and his daughters were unable to extract any meaning from them despite adequate decoding, because they had no linguistic comprehension in the relevant languages. Thus, the product takes into account that reading cannot be achieved without some element of both decoding and linguistic comprehension.

The combination of decoding and linguistic comprehension has been shown to account for at least three quarters of the variance observed in reading comprehension skill (Hoover & Gough, 1990). The Simple View is supported by the presence of double dissociations seen in reading-disabled populations. For example, children with dyslexia present with impaired decoding but preserved linguistic comprehension (e.g. Catts & Kamhi, 2005) whereas children with specific comprehension difficulties (poor comprehenders) present with impaired linguistic comprehension with decoding skills intact (e.g. Cain, Oakhill & Bryant, 2000; Stothard & Hulme, 1995).

In a longitudinal investigation of 182 eighth grade readers, Catts, Adlof and Ellis Weismer (2006) compared phonological, language and cognitive abilities of typical readers, poor readers and poor comprehenders. They found that poor comprehenders' phonological abilities were normal, but their reading comprehension assessed using three standard assessment instruments was impaired, whereas the opposite profile was seen in the poor readers. Retrospective analysis of their performance in tests administered when they were in kindergarten, second and fourth

grades showed that this dissociation was apparent at all time points. Thus, longitudinal evidence provides support for classifying reading disorders based upon the Simple View, (see Figure 1)



*Figure 1.1.* Classification of reading disorders based upon the Simple View of Reading

Behaviour genetic work provides further support for the Simple View of Reading. Keenan, Betjemann, Wadsworth, DeFries and Olson (2006) explored the genetic influences on composite measures of reading and listening comprehension. They measured reading comprehension and listening comprehension in 70 monozygotic and 121 dizygotic twin pairs. Analysis using Cholesky decomposition showed that genes had a strong independent influence on word recognition and listening comprehension. Both of these genetic components were found to exert independent influence on reading comprehension, and, crucially, they accounted for

all the genetic influence on reading comprehension skills. These results support phenotypic observations of comprehension difficulties that exist in the absence of word recognition difficulties. However, it must be remembered that genetic predispositions are modulated by environmental influences.

In summary, converging evidence supports the Simple View of Reading (Gough and Tunmer, 1986) as a framework for the classification of reading profiles. The studies reported in this thesis adopt this framework in the investigation of how the generation of different types of inference during reading may be influenced by decoding and linguistic comprehension.

#### *1.4 Children with Poor Comprehension Skills*

According to the Simple View, comprehension difficulty occurs when linguistic comprehension is compromised in the presence of adequate word identification.

In a longitudinal study Catts, Hogan and Adlof (2005) classified poor readers in accordance with the Simple View of Reading. In second grade, 15% fulfilled criteria for specific comprehension difficulties. However by fourth grade, 31% poor readers and 30% eighth grade poor readers were classified as having specific comprehension difficulty. This did not reflect shifts between classifications of reading disability. The comprehension difficulties were only becoming evident after decoding reached a reasonable level. The transition from ‘learning to read’ to ‘reading to learn’ in later stages of primary education is thought to lead to what is referred to as the “fourth-grade slump” (Chall, 1983; Catts, Hogan & Adlof, 2005). According to Nation (2007), approximately 10% of all children aged 7-10 years have difficulties with linguistic comprehension in the absence of word recognition impairments. Throughout this review such children are referred to as ‘poor comprehenders’.



Given that good comprehension is the product of a wide range of component factors working effectively together, it follows that a deficit in any one of those component factors has the potential to lead to comprehension failure. In general, poor comprehenders perform poorly on a range of tasks tapping component skills, indicating impairments that may relate causally, consequentially or incidentally to their comprehension difficulty (Cornoldi, de Beni, & Pazzaglia, 1996).

#### *1.4.1 General Cognitive Ability in Poor Comprehenders*

Nation, Clarke and Snowling (2002) explored the relationship between general cognitive ability and comprehension skill. Previous research suggested that poor comprehenders had depressed verbal ability relative to age-matched controls, (Stothard & Hulme, 1996) and this study replicated this finding in a sample of twenty-five 8-9 year old poor comprehenders. Overall, general cognitive ability was lower in the poor comprehenders, and this was accounted for by differences in verbal ability. Nation et al. also identified a subgroup of poor comprehenders with overall general cognitive ability below the normal range (<85). These children were not so much poor comprehenders as excellent decoders, and their comprehension skills were broadly in line with what their overall ability level would predict.

Across the poor comprehender group, spatial ability was equivalent to that of controls. Observed differences in general cognitive ability were accounted for by lower verbal ability and to a lesser extent lower non-verbal ability. In the average ability poor comprehenders, there was a large difference between verbal and spatial ability (spatial ability typically 13.5 standard score points higher than verbal ability), but in the low ability poor comprehenders the mean difference was smaller (5.4 standard score points difference in favour of spatial ability).

Studies differ with regard to whether they have reported poor verbal ability in poor comprehenders relative to controls matched on non-verbal ability (e.g. Nation, Clarke, Marshall & Durand, 2004; Nation, Snowling & Clarke, 2005; Stothard & Hulme, 1995; cf. Cain and Oakhill (2006a). The picture is complicated because of the varied selection criteria preferred by different investigators. In general, however, the evidence suggests that verbal skills are weaker in poor comprehenders than reading accuracy matched controls, but selection procedures may lead to the recruitment of children with either surprisingly impaired comprehension, or better reading accuracy than one might predict based upon their general ability in the presence of expected levels of comprehension.

#### *1.4.2 Vocabulary in Poor Comprehenders*

In a 2-year longitudinal study of 90 typically developing beginning readers, Muter, Hulme, Snowling and Stevenson (2004) found that vocabulary knowledge at age 4 years 9 months was a significant predictor of reading comprehension but not word recognition two years later. Similarly, in samples of older children, vocabulary has been shown to predict a proportion of the variance in reading comprehension (e.g. Cain, Oakhill & Bryant, 2004; Goff, Pratt & Ong, 2005; Seigneuric, Ehrlich, Oakhill & Yuill, 2000). The link between vocabulary and reading comprehension is complicated because it accounts for variance in both word reading and text comprehension (Cain & Oakhill, 2006a). If one adopts a definition of reading comprehension in accordance with the simple view, this accounts for some of the variance shared between word recognition and linguistic comprehension.

In some studies, poor comprehenders selected based on a discrepancy between their reading accuracy and reading comprehension show average levels of vocabulary (e.g. Nation & Snowling, 1998b), however in other studies vocabulary is found to be

weaker in poor comprehenders than in reading-age matched controls (e.g. Nation, Clarke and Snowling 2002). Notwithstanding this, Yuill and Oakhill (1991) found that children matched on vocabulary and word recognition differed in reading comprehension performance.

Cain (1994) suggested that vocabulary test selection might be important. In line with this, Daneman (1988) suggested that the relationship between vocabulary and comprehension test results reflect similarities in test requirements in that both measure information acquisition from context.

In order to explore the relationship between vocabulary and reading comprehension further, Cain, Oakhill and Lemmon (2004) divided 9-10 year old poor comprehenders into two groups depending on their vocabulary test scores. They found that poor comprehenders with weaker vocabulary skills benefited less from direct vocabulary instruction, requiring more repetitions to learn a novel word meaning. In support of this, Nation, Snowling & Clarke (2007) found poor comprehenders to be as good as chronological age- and decoding-matched controls at learning new phonological forms, but they required more direct instruction to learn their definitions.

Cain, Oakhill and Elbro (2003) found poor comprehenders had greater difficulty learning new words from context, and that this problem was further exacerbated if the words and information required to establish their meaning were separated in the text. Ricketts, Bishop and Nation (2008) found that poor comprehenders' ability to map orthographic forms to phonology was unimpaired relative to controls matched on chronological age and decoding ability. However, their semantic learning was impaired, particularly when tested after a delay. Interestingly, contrary to prediction, and despite showing impaired exception word

reading, the poor comprehenders did not show a proportionally greater deficit in orthographic or semantic learning of inconsistent novel word forms.

In summary, the precise relationship between comprehension skill and vocabulary remains to be clarified. Vocabulary is depressed in some groups of poor comprehenders, yet others selected using similar criteria show vocabulary levels equivalent to controls matched on chronological age and decoding ability. Studies of vocabulary acquisition consistently show that poor comprehenders have difficulty building associations between semantic and phonological forms relative to controls. This has been shown when they have been required to infer meanings from context and when they have been explicitly taught new words and their meanings. It is plausible, therefore, that the relationship between vocabulary and reading comprehension is mediated by inferential skills.

#### *1.4.3 Working Memory in Poor Comprehenders*

It has been suggested that a selective impairment in verbal working memory could be a common feature of those with difficulties in reading comprehension (Swanson & Berninger, 1995). Yuill, Oakhill and Parkin (1989) proposed a more general working memory impairment, as their sample of poor comprehenders performed more poorly on a counting span task as well as a listening span task. However others have failed to replicate this (Cain & Oakhill, 2006a; Stothard & Hulme, 1995). Cain and Oakhill (2006a) found that 7-8 year old poor comprehenders performed as well as good comprehenders on the counting span task, but performance on a listening span task was impaired.

Nation, Adams, Bowyer-Crane and Snowling (1999) point out that verbal skills are required for reading and recall of the digits, therefore there may be a verbal component to the counting span task. Furthermore, Nation et al. (1999) found a

selective impairment in verbal working memory in poor comprehenders, in the presence of normal performance on a spatial working memory task. Cataldo and Oakhill (2000) found no impairment in spatial working memory in poor comprehenders, lending further support to the notion of impairment specific to the verbal domain.

Studies reporting comparisons between good and poor comprehenders on tasks involving anomaly resolution or detection suggest that poor comprehenders are less able to detect or resolve anomalies when they are a distance apart in the text. When they are adjacent to each other, poor comprehenders' performance is much closer to that of controls (Yuill, Oakhill & Parkin, 1989; Oakhill, Hartt & Samols, 2005).

In summary, there does appear to be agreement in the literature that poor comprehenders do badly in tests of verbal working memory. However, there are three possible explanations for the differences in performance on working memory tasks shown by poor comprehender groups. Firstly, selection protocols using tasks with higher memory load may favour selection of poor comprehenders that perform relatively more poorly on working memory tasks; secondly, comprehension may be inhibited in the presence of poor working memory, and thirdly, as proposed earlier, poor performance in tasks tapping verbal working memory may reflect underlying language difficulties in poor comprehenders (Goff, Pratt & Ong, 2005; Nation et al, 1999).

#### *1.4.4 Language Skills in Poor Comprehenders*

Nation, Clarke, Marshall and Durand (2004) found an extensive range of oral language difficulties in a sample of poor comprehenders. They conducted a comprehensive assessment of reading and language skills in twenty-three poor comprehenders and twenty-two controls matched on chronological age, nonword

reading and non-verbal ability. The children were given tests of phonological processing, semantic processing, morphosyntax, and broader language skills. The poor comprehenders performed as well as controls on phonological tasks, however their performance in all of the language tasks was poorer than that of the controls. In approximately 30% of the poor comprehenders, the language difficulties were sufficiently severe to warrant a diagnosis of language impairment. Although there was significant overlap between the test scores obtained by these poor comprehenders and those observed in groups with specific language impairment (SLI), Nation et al suggest that there are qualitative differences in errors made by the two groups. For example, in past tense production, poor comprehenders make regularisation errors, whereas children with SLI are more prone to omitting grammatical markers.

Marshall and Nation (2003) found that poor comprehenders performed more poorly than controls on a sentence repetition task. Error analysis revealed a difference in the nature of their errors compared to those made by controls. Poor comprehenders made more errors, but also their errors were not in keeping with the gist of the passage. Good comprehenders typically reproduced the meaning of the sentence appropriately, even if recall of the words was inaccurate.

In line with this observation, semantic and syntactic deficits have been observed in poor comprehenders in a series of studies carried out by Nation, Snowling and colleagues (Nation & Snowling, 1998a, 1998b, 1999, 2000). They found that poor comprehenders' word naming latency was not facilitated by prior context to the same extent as that of controls (Nation & Snowling, 1998b). Furthermore, poor comprehenders showed deficits in semantic priming (e.g. Animal primes Dog), with functional priming (e.g. Shampoo primes Hair) preserved (Nation & Snowling, 1999). In a comparison of phonological and semantic skills, Nation and Snowling (1998a)

showed that poor comprehenders performed normally on a rhyme judgement task, but were poorer than controls in a synonym judgement task.

Nation and Snowling (1998a) also found poor comprehenders to be worse at reading exception and low frequency words, a finding recently replicated by Ricketts, Bishop and Nation (2008). This conforms to predictions that follow logically from the connectionist triangle model of reading developed by Plaut, McClelland, Seidenberg and Patterson (1996). The model suggests that a breakdown in semantic processing would lead to difficulty with the recognition of exception and low frequency words, exactly as Nation and Snowling (1998a) observed. In line with this, Goff, Pratt and Ong (2005) found that irregular word reading measured using a test of exception word reading was the strongest independent predictor of reading comprehension once age and cognitive ability had been controlled.

Thus, the evidence suggests that poor oral language skills are common in poor comprehenders. Returning to the issue of overlap between reading comprehension difficulties and SLI, it is important to note that a child's profile of reading and language ability may change over time. A series of longitudinal studies by Bishop and colleagues following the progress of 87 children with a diagnosis of SLI at age 4, (Bishop & Adams, 1990; Bishop & Edmondson, 1987; Stothard, Snowling, Bishop, Chipchase & Kaplan, 1998) has shown that the profile of difficulties exhibited by children with SLI changes over the course of development. When assessed at 5 years 6 months, 30 of these children no longer showed a profile consistent with SLI suggesting that the problem had now resolved, 38 children had persisting difficulties consistent with their SLI diagnosis at age 4, and 19 children were showing a profile of general developmental delay (Bishop & Edmondson, 1987). When their reading profile was assessed at 8 years 6 months, those with persistent SLI were shown to

have adequate reading accuracy, but scored relatively poorly on measures of reading comprehension. Therefore at this stage in development, such children could plausibly be identified as poor comprehenders.

Later work has found that children from both the persistent and resolved SLI groups subsequently developed reading difficulties affecting both phonology and linguistic comprehension at age 15-16 (Stothard et al, 1998), indicating that any apparent recovery shown by these children in the early years was not maintained.

In another study looking at this sample, Snowling, Bishop and Stothard (2000) showed that good performance IQ protected against some of the deleterious effects of language impairment on subsequent literacy development. They found that for those children with performance IQ greater than 100, their reading accuracy and comprehension declined to a lesser extent between the ages of 8 and 15 years, than those with performance IQ less than 100.

In summary, there are areas where reading comprehension impairment and language impairment may overlap, to the extent where poor comprehender groups may include children with underlying language impairment severe enough to warrant a diagnosis of SLI. The heterogeneity seen in samples of poor comprehenders and language-impaired children complicates the picture further. Recent longitudinal studies tracking early development in children who subsequently develop reading comprehension difficulties indicate that oral language weaknesses shown by poor comprehenders precede the development of reading comprehension impairments, however precise causality has not yet been established (e.g. Nation, Cocksey, Taylor & Bishop (2010 in press.)), and these are discussed more fully later in this review.



#### *1.4.5 Poor Comprehenders' Understanding of Story Structure*

Cragg and Nation (2006) found structural limitations in stories written by a small sample of poor comprehenders. Eleven poor comprehenders and 19 typically developing children were asked to write a story based on a series of pictures provided. After a period of time had passed, the children were also asked to re-tell the story orally. The written stories produced by the control group were rated as more sophisticated in terms of story structure, and contained more main ideas than those produced by the poor comprehenders. Unlike the typically developing children, the poor comprehenders did not present more main ideas when producing the story orally. Moreover, the structure of the oral stories produced by the poor comprehenders was less sophisticated. As the studies reviewed require story production, although results are consistent with poor understanding of story structure, they could be more indicative of expressive language difficulties or memory problems.

Interestingly, poor comprehenders' story production can be improved simply by providing a more directive story title (Cain, 2003). This finding suggests that poor comprehenders may not activate their knowledge of structural conventions spontaneously, but can do so if prompted. Therefore it seems that poor comprehenders' knowledge of story is intact, but their awareness of when to apply this knowledge may be lacking.

#### *1.4.6 Comprehension Monitoring in Poor Comprehenders*

Garner and Kraus (1984) compared comprehension monitoring in 12-13 year old good- and poor comprehenders using an error detection task. Their results suggested that poor comprehenders were less able to detect inconsistencies in passages both within and between sentences, however in this study poor comprehenders were not selected using an objective measure of comprehension performance. Using

objectively selected groups, Yuill, Oakhill and Parkin (1989) found that poor comprehenders performed as well as good comprehenders in an anomaly resolution task when the anomaly and resolving information were adjacent in the text. However, poor comprehenders were worse at detecting anomalies if the relevant pieces of information were several sentences apart in the passage. These results are corroborated by previous work showing similar effects of distance on anaphor resolution (e.g. Yuill, Oakhill & Parkin, 1989), and on the ability to detect sentence-level inconsistencies (Oakhill, Hartt & Samols, 2005). More recently, a positive relationship has been demonstrated between comprehension monitoring behaviour and comprehension ability that is independent of working memory (Cain, Oakhill & Bryant, 2004).

Using a cross-sectional design, Hacker (1997) studied monitoring activity in 315 secondary school children in 7<sup>th</sup>, 9<sup>th</sup> and 11<sup>th</sup> grades. He asked participants to search through a text three times. On the first pass, it was found that better readers and older children were more likely to detect errors at the lexical, syntactic and semantic levels. On the second reading, after being explicitly directed to look for meaning errors, further differences between groups emerged, with little improvement in error detection performance in the less skilled readers of all three age groups. It seems that poor comprehension monitoring is consistently associated with poor reading comprehension.

In summary, accuracy and frequency of comprehension monitoring is positively related to reading comprehension ability. Studies of poor comprehenders have found evidence of impaired comprehension monitoring in every age group tested. However, it is also the case that a minority of poor comprehenders show intact comprehension monitoring (Cain & Oakhill, 2006a).

#### 1.4.7 Identification of a core deficit in poor comprehenders

Impairments in all of the basic and higher level skills required for linguistic comprehension have been observed in poor comprehenders, however studies in general fail to disentangle the causal relationships between the various deficits. Furthermore, there are inconsistencies in the literature, with some studies showing a relationship between a given variable and reading comprehension level, and others refuting it. As yet there is no consensus regarding a “core” deficit common to all children experiencing comprehension difficulty (cf. phonological skills in dyslexia), and there is strong evidence to suggest that poor comprehenders are not an homogenous group. It is also plausible that selection protocols may favour poor comprehenders with particular profiles, and some of the heterogeneity in findings observed between and within studies could be attributable to differences in selection criteria preferred by different researchers.

A recent longitudinal study by Nation, Cocksey, Taylor & Bishop (2010 in press), found evidence of weaker nonphonological language skills in 8 year old poor comprehenders compared with reading and chronological age-matched controls. Retrospective analysis of data collected from these children at previous timepoints (at ages 5, 5.5, 6 and 7 years showed that these mild deficits in understanding of grammar, receptive and expressive language, and listening comprehension were evident throughout. Importantly, this shows that the oral language weaknesses seen in poor comprehenders cannot be a consequence of a reading comprehension deficit. It is plausible, however that they may be causally related to the development of reading comprehension problems. In line with the Simple View of Reading, these children showed normal phonological skills at all timepoints.

These results were also consistent with other studies that have demonstrated poor nonphonological skills in poor comprehenders (e.g. Catts et al, 2006) and studies showing that preschool nonphonological language skills predict subsequent reading comprehension skills (eg. Kendeou, van den Broek, White & Lynch (2009).

Cain and Oakhill (2006a) suggest that low level component skills such as vocabulary, interact to impact upon performance in higher-level skills such as comprehension monitoring, inference generation and understanding story structure. Following on from this, it is possible that not all poor comprehenders will show a deficit in a given low-level skill such as vocabulary, but an individual's profile across a variety of component skills will facilitate or constrain their reading comprehension.

To summarise, recent evidence from longitudinal studies suggests that nonphonological language deficits are present in children with reading comprehension difficulties, and these difficulties precede the development of those reading comprehension difficulties. However they have failed to identify a single deficit that is present in all poor comprehenders, rather, in support of the Simple View of Reading, the findings suggest that oral language weaknesses that impact upon linguistic comprehension are associated with reading comprehension deficits, and within these, the profile of such weaknesses can differ significantly between affected individuals. In line with this, a recent large scale randomised controlled trial of three different interventions targeting reading comprehension, found that reading comprehension skills of 8 year old poor comprehenders were best improved by an intervention targeting oral language skills. This intervention programme was found to yield better results in terms of improvements in reading comprehension skill than an intervention similar in design targeting text comprehension, and an intervention targeting both text comprehension and oral language skills. Furthermore, Clarke et al

did not identify a single component skill that was impaired to a similar degree in all poor comprehenders.

Thus, the evidence available to date does not support the view that a core deficit in a single component skill leads to reading comprehension difficulties. Rather, it suggests that broad based but relatively mild weaknesses in nonphonological language skills are implicated in the subsequent development of reading comprehension deficits. For this reason, the Simple View of Reading is used as a framework for the experiments in this thesis, which will go on to compare performance of poor comprehenders with that of typically developing readers and poor decoders.

The remainder of this review focuses in more detail on inference generation, and its relationship with reading comprehension. Firstly, the range of inference types that readers can generate during reading will be discussed, and their position in relation to theories of inference generation will be explored.

### *1.5 Inference Generation and Reading Comprehension*

Earlier in this review, the role of inferences in successful reading comprehension was outlined. It is clear that in order to construct a sufficiently intricate situation model for adequate comprehension, inferences need to be made to resolve gaps in coherence. These gaps can be either at a local level or a global level. At the local level, the information enabling the reader to generate the inference that fills the gap is close by, such as in the example below,

*The dog saw a cat. He barked and pulled hard on the lead.*

At the global level, information encountered much earlier in the text must be accessible in order for the appropriate inference to be generated. The nature of

language is such that readers need to be able to make a diverse range of inference types integrating many different pieces of information.

### *1.5.1 Classification of Inference Types*

There are two main dimensions according to which inferences are classified. Firstly inferences can be classified as coherence inferences or elaborative inferences. Coherence inferences are those that must be made for a text to make sense. These can involve the resolution of cohesive devices such as anaphora, or bridging inferences made to fill information gaps in the text, via the application of real world knowledge. There is some consensus that resolution of cohesive devices occurs at the textbase level of comprehension, whereas bridging inferences are made during construction of the situation model (e.g. Graesser, Singer & Trabasso, 1994). In the case of elaborative inference, although the passage can be understood without the inference, the mental representation of the text and the resulting situation model will be more detailed when the inference is made. Elaborative inferences therefore build on the richness of the situation model, and they tend to involve the utilization of the reader's knowledge base in order to make predictions or assumptions.

Secondly, Kintsch and Rawson (2007) suggest that inferences can be classed as knowledge-based or text-based inference. In order to generate some inference types, the reader needs to call on background knowledge in order to fill the gap. For example, consider the sentence:

*“Tom put the milk in the refrigerator. He closed the door.”*

Here, the reader needs to infer that the door was that of the refrigerator in order for the passage to be coherent. In order to do that, the reader must be aware that refrigerators have doors, and that the specific door that is being referred to is that of the refrigerator.

An example of text-based inference would be something like:

*“Emma sings louder than Sarah. Catherine sings loudest of all.”*

From this information alone it can be inferred that Catherine has a louder voice than Emma, and there is no need to refer to background knowledge.

At a finer grained level, inferences can be classified in terms of the knowledge or knowledge-type that is required in order to generate them, such as inferences about time, spatial dimensions or emotional states; inferences relating to semantic, functional or causal relations; object properties or typical behaviours; they can also be classified in terms of the referent type such as anaphoric references or pronouns.

#### *1.5.2 Measurement of Inference Generation*

Researchers have employed numerous techniques to uncover mechanisms underpinning the generation of inferences. For many years it was common for inferential processing to be tested using off-line measures, with the generation of a particular inference being assessed retrospectively. More recently, on-line behavioural measures have been used to measure some aspects of inferencing behaviour as it occurs in real-time.

##### *1.5.2.1 Off-line measures of Inference Generation*

In many early studies participants were primarily assessed on their ability to answer questions presented in a variety of response formats (e.g. multiple choice, free response, dichotomised), with a correct response requiring either literal information from the passage to be recognised or recalled, or for some type of inference to be made. In more recent comprehension research, including the work presented in this thesis, comprehension questions have been employed as a secondary measure, usually to ensure that participants are reading for meaning or to see whether participants are able to generate particular inference types. They are also widely used

in standardised tests that measure comprehension skills and therefore participants may be selected on the basis of their ability to answer comprehension questions.

Other studies have used off-line techniques such as testing accuracy of recall and recognition of material presented as an index of inferential processing (e.g. Bransford, Barclay and Franks, 1972; Oakhill, 1982, Paris & Lindauer, 1976). However given the technology available and the increased information that may be obtained by looking at inference generation as it happens, on-line comprehension measures are now the gold standard in inference research.

#### *1.5.2 2 On-line Measures of Inference Generation*

An on-line measure allows us to measure and monitor behaviour in real time. In the context of inference generation, on-line measures enable the study of characteristics of specific inferences made including their time-course, whether they are made automatically and the grain of the information that is actually inferred (for example, do we infer that a character is feeling a particular emotion-state, or do we infer a range of emotion-states that may be plausible in a given situation?) They can also be used to examine the conditions required for the generation of a particular inference type to be triggered or prevented. Examples of on-line measures include tests embedded at strategic points within the text such as lexical decision tasks and naming speed measures (e.g. Nation & Snowling, 1999, 1998a, 1998b), sentence verification tasks where speed of participants' responses to true/false or yes/no dichotomies enable inferences generated on-line to be distinguished from those generated post-hoc (e.g. Bowyer-Crane, 2002; Myers, O'Brien, Albrecht & Mason, 1994 experiment 4); eye-tracking measures whereby readers' eye movements are plotted and used to specify when particular inferences are generated (e.g. Nation, Marshall & Altmann (2003)); and timing measures whereby reading speed is



monitored and comprehension difficulties reflecting inference generation and integration are represented by inhibition of reading in the critical area of the passage. Measurement of brain activity during comprehension task performance using functional neuroimaging techniques such as whole head magnetoencephalography (MEG) represents an exciting new development that has yet to be fully exploited in comprehension research.

### *1.5.3 The Time Course of Inference Generation*

As Kintsch and Rawson (2007) observe, there is ongoing debate about the time course of the generation of different types of inference, and which inferences are generated automatically during reading. The constructionist view (e.g. Graesser, Singer & Trabasso, 1994) mentioned earlier suggests that the reader makes all possible inferences during reading, or on-line and new information is mapped on to a single comprehensive situation model. Minimalists would argue that the reader makes as few inferences as they can get away with. These arguments relate to what readers accept as a standard for coherence. Constructivist theory advocates that when reading a passage, the reader strives to maintain both global and local coherence. That is, in order for reading to proceed smoothly, the portion of the text in current focus must be consistent with the relevant information delivered in all earlier parts of the text. According to the minimalist viewpoint, global coherence is not maintained, and only breaks in local coherence would inhibit reading.

In a series of studies O'Brien, Albrecht and colleagues used measurement of reading times to uncover information about how readers amalgamate incoming material with existing representations, whether they strive for global or local coherence during reading and inference types that are generated on-line. They also explored factors that inhibit a previously generated inference. In all of these

experiments participants were asked to read short fictional narratives presented on a computer screen sentence by sentence, and had to press a “line advance” key to move from one sentence to the next. The reading time measure obtained was the time between presses of the line advance key. Participants were told that the most important component of the experimental task was to answer a single comprehension question presented at the end of each passage accurately and as rapidly as possible. In some passages readers were presented with a target sentence that contained an inconsistency that either led to a break in global or local coherence. Reading times to inconsistent sentences were compared with reading times to the same sentences presented in text that meant that they were consistent with context.

For example O’Brien and Albrecht (1992) asked participants to read the following story (target sentence shown in italics).

“As Kim stood (inside/outside) the health club she felt a little sluggish. *She decided to go outside and stretch her legs a little.* She was getting anxious to start and was glad when she saw the instructor go in the door of the club. Kim really liked her instructor. Her energy and enthusiasm were contagious.”

In this story, the consistency of the target sentence was modified by altering whether Kim was inside or outside the health club. If Kim was outside the health club, saying that she decided to go outside would be inconsistent with the context. If text is monitored for violations in local coherence, then reading times to the target sentence in the inconsistent condition will be increased. A further manipulation was included to determine whether global coherence is also monitored. Half of the participants read this story with three sentences of filler information placed before the target sentence. If, as the constructivist position indicates, readers strive to maintain both local and global coherence, then reading times of all target sentences in the

inconsistent condition will be increased as readers attempt to solve the anomaly. However, if the minimalist view prevails, then only inconsistencies in the local condition will affect reading times.

In line with the constructivist view, O'Brien and Albrecht (1992) showed that inconsistencies about a story character's location that break either global or local coherence lead to inhibition of reading from the point at which the inconsistency is encountered. Albrecht and O'Brien (1993) adapted this paradigm to explore how new information mapped onto existing representations. Where narratives contained breaks in coherence at the global, but not local level, they found reading of (inconsistent) target sentences was inhibited. They argued that, in line with constructivist theory, this provided strong evidence that incoming information was being contemporaneously integrated into a single comprehensive situation model. In further support of this claim, participants showed enhanced recall of passage details up to and including the inconsistent sentence, suggesting that reprocessing of this part of the passage had occurred. This recall was superior to that measured to wholly consistent passages.

Other studies using this paradigm lend support to the constructivist view (e.g. Albrecht & Myers, 1995; Myers, O'Brien, Albrecht & Mason, 1994) suggesting that a wide range of inferences are generated on-line, and that features of the text can manipulate the magnitude of the inhibition caused by inconsistency with context.

#### *1.5.4 Summary*

In summary, the measurement of reading time has been successfully used in previous experiments to examine standards for coherence and mapping of incoming material with existing representations. In order for a consistency effect to be shown, the reader must have inferred something on-line about the protagonist based on earlier

portions of the text. The experiments reported above tap into different types of inference, and show that once an inference has been made, its trace is stored in memory during reading, and reactivated when the current text has some overlap with it. The results also suggest that in general, typical readers strive to maintain global and local coherence, in line with the constructionist position. From a methodological perspective, they also highlight how textual features can moderate levels of inhibition shown when an inconsistency is encountered.

### *1.6 Children's Inference Generation*

Evidence suggests that prior to reading acquisition, children make inferences in order to understand sequences of events that are similar to those made during reading (Van den Broek et al, 2005). As reading is acquired, inference generation develops between the ages of six and ten years (Paris & Upton, 1976). This ability to generate inferences is consistent for language processing in the written, oral and audio-visual domains (Kendeou, Bohn-Gettler, White & van den Broek, 2008). According to van den Broek et al (2005), a developmental sequence in inference types that children can generate, shown in table 1, is seen as children mature. It is suggested that text complexity relative to decoding level, rather than ability to generate inferences per se may drive this sequence and constrain generation of more complex inferences in emergent readers.

*Table 1.1.* Developmental sequence of inference types in narrative comprehension (reproduced from van den Broek et al, 2005).

Order	Inference type
1	Concrete physical relations that occur close together
2	Concrete physical relations between distant events
3	Causal relations involving character's goals, emotions and desires
4	Hierarchical and thematic relations between clusters of events
5	Translation of story theme into moral or lesson

#### *1.6.1 Inference Generation in Poor Comprehenders*

Research by Oakhill and colleagues has consistently shown that poor comprehenders make fewer inferences than typically developing readers matched on chronological and reading age (e.g. Yuill & Oakhill, 1991; Cain & Oakhill, 1999, Cain, Oakhill, Barnes & Bryant, 2001). This seems to be consistent across a range of inference types such as anaphor and pronoun resolution (Yuill & Oakhill, 1986, 1988a).

Oakhill (1982) found that poor comprehenders were more accurate at correctly identifying sentences they had read before than typically developing readers. Typical readers often indicated incorrectly that they had read a sentence before when the meaning could have been inferred from the text that they had actually read. It was suggested that poor comprehenders did not (falsely) recognize these sentences because they had not inferred their content, but they were able to identify the words that they had seen before.

In line with this, Oakhill (1983) found that poor comprehenders did not instantiate meanings like typical readers, and that this was not related to deficits in

knowledge. For example, in typical readers, recall of a sentence like “the fish attacked the swimmer”, was facilitated by the word “shark” indicating that they had activated the word shark when reading the sentence. In poor comprehenders, presentation of instantiated nouns did not cue recall.

It has been demonstrated that knowledge availability does not explain the failure to generate inferences shown by poor comprehenders, as poor comprehenders make fewer cohesive and elaborative inferences than good comprehenders even when the knowledge base is strictly controlled and taught until they demonstrate perfect recall (e.g. Cain, Oakhill, Barnes and Bryant, 2001).

Cain and Oakhill (1999) explored inferential skills of poor comprehenders compared with two control groups, typical readers matched on both chronological and reading age and a comprehension age-matched group of younger children. All groups read a set of stories followed by six comprehension questions. Two questions required application of literal knowledge of the passages, two required generation of text-connecting inferences and two required generation of gap-filling inferences. Poor comprehenders answered a similar number of literal questions correctly, but were worse than both control groups at answering text-connecting inference questions. They were worse than the typical readers at answering gap-filling inference questions, but performed at a similar level to the comprehension-age matched group. Overall, they concluded that poor comprehenders’ inferential skills were worse than those of both control groups suggesting that this deficit may be causally implicated in their comprehension difficulty. For the gap filling inferences, there was a trend for poor comprehenders to be worse than the comprehension-age matched group, however, as the authors argue, in such a young sample it is plausible that the ability to generate gap-filling inferences was not fully developed.

Cain and Oakhill's (1999) findings also showed that these inferencing difficulties do not reflect poor memory for text and deficits in knowledge. Poor comprehenders were able to answer as many literal questions as typical readers, and further questioning showed that they had the general knowledge required to generate the inferences that they failed to make spontaneously. It was therefore suggested that less skilled comprehenders lack awareness of how or when to apply their general knowledge to make inferences, and therefore despite being able to make inferences, they do not make as many inferences automatically as typical readers.

Oakhill (1984) compared inferential and memory skills in 12 poor comprehenders and 12 controls matched on chronological age, reading age, gender and vocabulary. Participants read four short stories, and were told they would be asked some questions about each story. After reading each story they were asked eight open comprehension questions with the text removed. The questions were repeated with the passage available. For each passage, four comprehension questions required literal information, and four required generation of an inference. When the passage was not available, poor comprehenders were worse at answering both literal and inferential questions but their performance on the literal questions was equivalent to controls when the passage was available. The poor comprehenders did not, however, benefit from the presence of the text when required to answer inference questions. However, Cain and Oakhill (1999) showed that prompting to use the text facilitated their ability to answer some, but not all, inferential questions.

Yuill and Oakhill (1988) evaluated the effects of three training programs on three groups of poor comprehenders and controls. They found that a training program focusing specifically on inferences led to similar gains in comprehension to training in overall comprehension, and that a program involving rapid decoding practice did not

impact on comprehension. They found that training poor comprehenders to generate more inferences led to overall improvements in comprehension, as did the more general comprehension training program. It is unclear whether the training taught them strategies to overcome their difficulties, or taught them to read for understanding in the same way that skilled comprehenders do, as controls showed no significant improvements in comprehension as a result of these programs. Training in rapid decoding had no effect on either poor comprehenders' or controls' performance in reading accuracy or comprehension. This fits with the Simple View of Reading, suggesting that poor comprehenders do not have problems with decoding.

### *1.6.2 Summary*

Taken together, these studies show that poor comprehenders make fewer inferences than typically developing children when they read. Results of studies using comprehension age-matched designs suggest that this may be causally implicated in their comprehension difficulty, and the results of inference training lend support to this view.

The published evidence shows that poor comprehenders' inferencing difficulties do not reflect an inability to actually generate particular types of inference, rather that they do not seem to engage in as much spontaneous inferential activity. It is plausible, therefore, that poor comprehenders' failure to generate inferences could arise because they read text in a qualitatively different way from skilled comprehenders, in that reading text proceeds in a more passive manner. In this way, it could be argued that poor comprehenders generate inferences in a more minimalist fashion than typically developing readers, and in contrast with the results of reading time experiments in adults outlined above that show results that support Kintsch's constructivist viewpoint (Kintsch, 1998). Off-line measures also suggest that they do



not generate the same number of inferences as controls, and fail to use relevant strategies or cues to bootstrap their understanding.

### *1.7 Aims of this thesis*

The experiments reported in this thesis focus on one specific aspect of linguistic comprehension, inference generation. Performance of 7-10 year old typical readers is compared with that of age-matched poor decoders and poor comprehenders, to determine whether, in line with predictions based on the framework suggested by the Simple View of Reading, dissociation is observed between linguistic comprehension and word recognition.

In Experiments 1-4 reading time is used as the primary measure of inference generation. This methodology was selected because first, it has been shown to be a useful and robust measure of on-line inferencing behaviour in adults, and second it is suitable for use with developing readers, providing the texts are simple enough for them to decode.

In line with methodology used by Gernsbacher and colleagues to investigate emotion-state inference processing in adults, (Gernsbacher, Goldsmith & Robertson, 1992; Gernsbacher & Robertson, 1992; Gernsbacher, Hallada & Robertson, 1998), a proportion of the test passages contain inconsistencies that contradict information a reader could reasonably be expected to infer from the preceding text. In typical adult readers, such inconsistencies lead to slower reading, as they detect the inconsistency with inferred context and then attempt to resolve it. However, reading will only be slowed if an inference such as an assumption about the protagonist's emotion-state has been made on-line during reading, because when information is read that contradicts the 'expected' emotion-state or range of emotion-states, they perceive this as an inconsistency that needs to be resolved. In this series of experiments, inference types

and textual factors are manipulated to examine factors that may constrain or facilitate inference processing. Comprehension questions are asked at the end of each test passage to check that participants were reading for meaning, and as an off-line additional measure of their inferential skills.

It is predicted in Experiments 1-4 that poor comprehenders will show less inferential processing than either typical readers or poor decoders, evidenced by less disruption of reading when inconsistencies are encountered in the text. It is also predicted that poor decoders will show intact inferential processing, as long as decoding is achieved.

In Experiment 5, an analogous procedure is followed in the aural domain, to determine whether the findings of Experiments 1-4 are mirrored for aurally presented texts. It is predicted that if poor comprehenders' inferencing difficulties are specific to linguistic comprehension, a similar pattern of results will be obtained in the aural modality. In line with this, it is predicted that poor decoders will perform like typical readers, without being slowed down by effortful decoding.

In the final experiment reported in this thesis, one possible explanation for between-group performance differences is explored. It is plausible that poor comprehenders are either less likely to notice, or are more tolerant of ambiguities or inconsistencies in text because they adopt a lower standard for coherence when they read. This would logically drive them to engage in less inferential processing than typical readers, as their standard is met using fewer inferences. This study therefore is designed to test the hypothesis that poor comprehenders adopt and implement a lower standard for coherence than either poor decoders or typically developing children.

## CHAPTER TWO

### **Reading Disability and Children's On-line Generation of Emotion-state and Spatial Inferences**

#### *2.1 Introduction*

For typically developing readers, the improvement of decoding skills and increases in their automatisisation bring about increases in reading comprehension. Eventually, comprehension of text reaches similar levels to those of oral language comprehension (Gernsbacher, 1990; Curtis, 1980). The Simple View of Reading (Gough & Tunmer, 1986) suggests that decoding and linguistic comprehension exert independent influences on reading and its development. There is substantial empirical evidence supporting this view from a diverse range of sources including longitudinal and cross-sectional studies (e.g. Catts, Hogan & Fey, 2003) and, recently, behaviour genetic research (Keenan, Betjemann, Wadsworth, DeFries & Olson, 2006).

The Simple View provides a framework within which reading disability can be classified. In the case of poor comprehenders, decoding develops normally, but linguistic comprehension is impaired. Conversely, in poor decoders, decoding is impaired in the presence of intact linguistic comprehension. In the case of poor decoding, there is consensus that poor phonological skills lead to impaired development of grapheme-phoneme correspondences resulting in slow and laborious decoding (e.g. Snowling, 2000). However, to date an underlying core deficit common to all poor comprehenders has not yet been identified (Cain & Oakhill, 2006a).

Depressed levels of vocabulary (e.g. Nation & Snowling, 1998a), working memory (e.g. Yuill, Oakhill & Parkin, 1989), comprehension monitoring (e.g. Cain, Oakhill & Bryant, 2004), inference generation (e.g. Yuill & Oakhill, 1991), understanding of story structure (e.g. Cain, 2003), general cognitive ability (e.g.

Nation, Clarke & Snowling, 2002), or language ability (e.g. Nation, Clarke, Marshall, & Durand, 2004) have been found in some, but crucially not all, poor comprehenders.

In the case of inference generation, successful reading requires some inference types to be generated on-line during reading, in order that the reader can construct a sufficiently elaborate situation model to facilitate timely comprehension. Studies of children with reading comprehension difficulties have found that, although these children can make inferences, they do not do so as readily as typical readers (e.g. Bowyer-Crane, 2002; Oakhill, 1982; 1983; 1984; Yuill & Oakhill, 1991). This deficit is not considered to be accounted for by lack of knowledge availability or poor memory for text (e.g. Cain & Oakhill, 1998; Cain, Oakhill, Barnes & Bryant, 2001)

The majority of previous studies examining inference generation during reading have employed off-line (retrospective) measures, or tasks which require the flow of reading to be interrupted, such as by interspersing a rapid naming task to establish whether a given inference has been made (e.g. Gernsbacher & Robertson, 1992). In a series of experiments investigating emotion-state inferences made during reading by adults, Gernsbacher and colleagues used a self-paced reading paradigm where the reader was presented with short narratives sentence by sentence. The reader was required to press a key to advance to the next sentence in the story. This method has been used to measure inference generation on-line, comparing reading times (measured as time between key presses) to target sentences containing emotion-states consistent and inconsistent with context (e.g. Gernsbacher, Goldsmith & Robertson, 1992; Gernsbacher & Robertson, 1992; Gernsbacher, Hallada & Robertson, 1998).

In these experiments, participants were typically required to read a series of 7-10 sentence long narratives, in which a real-life situation was described, that would induce strong emotion in the main character. However, at no point prior to the last

sentence was the main character's emotion state explicitly mentioned. In this final sentence, the emotion of the main character was stated, and between-subjects comparisons were made between reading times recorded for sentences containing emotions consistent and inconsistent with context. These studies found that reading was slowed down (inhibited) when inconsistency with context was encountered. Furthermore, the inhibition suggests that prior to reading the emotion-state descriptor, an inference had been generated about the character's emotional state. Gernsbacher and colleagues used this paradigm to examine the breadth, temporal properties and automaticity of emotion-state inferences, and manipulated factors that might constrain inference processing such as semantic and attentional factors (Gernsbacher, Hallada & Robertson, 1998, Gygax, Oakhill & Garnham, 2003).

Gernsbacher, Goldsmith and Robertson (1992) compared reading times to target sentences containing emotions consistent and inconsistent with context. They also varied the semantic distance between the consistent and inconsistent emotion states. In experiment one, the inconsistent emotion states were perceived opposites of the consistent states (e.g. sad-joyful, guilty-proud), which were judged to be similar on other emotional state dimensions sharing the same intensity, duration, self-relevance and temporal reference (Frijda, 1986). In experiment two, the inconsistent emotional states were of the same affective valence (positive or negative), not opposites, and unlikely to be states that would be felt in the context described (e.g. bored-angry, grateful-confident). It was found that the magnitude of inhibition (reading time differences between consistent and inconsistent sentences) depended on the extent of the inconsistency. For example, inconsistency caused greater disruption when the emotion was a perceived opposite (e.g. happy-depressed) than when it was an inappropriate emotion-state sharing the same valence (e.g. happy-sympathetic).

Across both experiments, reading times to consistent sentences were remarkably similar.

Further studies using the same paradigm have replicated and extended these findings. Gernsbacher and Robertson (1992) showed that density of emotion content within the experimental stimuli affected reading times to inconsistent sentences. They found that in the low density condition (75% neutral stories, 25% emotional stories), inconsistent target sentences were read significantly more slowly than when the density of the emotion content was higher. It was suggested that in the high density condition (75% emotional stories, 25% neutral), activation to all emotion-states was increased. Alternatively, it could be suggested that the anomaly was more unexpected in the low density condition, and therefore reading was further disrupted.

Gernsbacher, Hallada and Robertson (1998) investigated the automaticity of emotion state inferences by interspersing tasks to divide the reader's attention. In experiment one this additional task was a simple tone identification task, in experiment two they chose a more challenging memory load task, and in the third experiment a cumulative memory load task was employed. Reading of target sentences was generally inhibited by the presence of the additional task, and this was more evident for the more demanding cumulative memory load task. However, the additional inhibition of reading generated by the inconsistency was unchanged. Divided attention tasks therefore do not compromise the activation of emotion knowledge, suggesting that this activation proceeds automatically.

Gygax, Oakhill and Garnham (2003) examined specificity of emotion-states inferred during reading adapting the stories created by Gernsbacher, Goldsmith & Robertson (1992). No reading time differences were found between a consistent synonym and consistent similar target emotion; however reading times were

significantly longer to target sentences containing an inconsistent emotion state as before. These results suggest that the inferences generated are not emotion-specific, and encompass a narrow range of emotional states that a character could plausibly be feeling. The authors go further, suggesting that readers only infer the most salient emotion subcomponents automatically during reading and therefore the inferences generated concern a range of properties of the emotional state of the character.

Soederberg and Stine (1995) investigated age-related changes in emotion state activation at discourse level. Performance of twenty-one young adults (mean age 20.4 years) and forty-three older adults (mean age 68.8 years) was compared. A robust main effect of consistency was shown when reading target sentences, and no difference in performance was observed between the two age groups tested. From the literature reviewed, it would appear that this is the only example of a study in which this paradigm has been used to measure between-group differences.

Together, this series of experiments demonstrates that inferences about fictional characters' emotional states are made during reading. The presence of an inconsistency inhibits reading proportional to the extent of the anomaly. Such slowing may reflect the creation of substructures to accommodate the anomaly, with larger mismatches requiring more elaborate restructuring of the mental representation. As the experiments illustrate, this paradigm can be used to uncover information about the nature of inferences generated about fictional characters' emotion-states during the reading of short narratives, at least in adults. The study presented here applies this technique to determine whether children generate emotion-state inferences on-line, and whether generation of these inferences depends upon reading ability. A set of new stories were constructed that were similar in structure to those used by Gernsbacher, but suitable for readers aged between seven and ten years. The

methodology was adapted so that within-subjects comparisons could be made, to control for individual differences in reading speed that are more likely to be observed in developing readers. In order to obtain a baseline measure of reading speed a neutral condition was introduced, to account for the possibility that reading of consistent sentences may be facilitated due to prior activation of emotion state words.

This study addresses two main research questions; Firstly, during the reading of short narratives, do children aged 7-10 years infer the emotion-state of the protagonist? Studies of typically developing readers indicate that they are able to generate inferences during reading (e.g. Yuill & Oakhill, 1991; Bowyer-Crane, 2002), and the inhibition of reading when there is an inconsistency, to date only shown in adults provides evidence of on-line inference generation. The first hypothesis therefore predicted that, like in adult readers, children would read inconsistent target sentences more slowly than either consistent or neutral target sentences

Second, how does reading difficulty affect inference generation in this age group? In order to test this, typically developing readers were compared with two reading disabled groups - poor decoders and poor comprehenders. It was hypothesized that, in line with poor comprehenders' previously documented problems with inference generation, they would exhibit less inhibition of reading when presented with an inconsistent sentence. In poor decoders, it is hypothesized that overall reading would be slower but they would show similar levels of inhibition to typical readers when presented with an inconsistency.

A third issue explored in this experiment was whether on-line inference generation leads to facilitation of reading when the text is coherent. If an inference about the protagonist's emotion-state has been generated, activation of the emotion-state label and related words may already be activated. Based on this idea, it was



hypothesized that the mean syllable reading speed to consistent sentences will be faster than syllable reading speed to neutral sentences matched on word length, sentence position in the narrative, and differing from the consistent sentences by no more than two words.

## *2.2 Experiment 1 Method*

### *2.2.1 Materials*

#### *2.2.1.1 Experimental Stories*

Five sets of three stories were used in the experiment, one set for each of the five basic emotions happiness, sadness, fear, anger and disgust. All three stories in a set described situations where the protagonist would be expected to feel the same emotion. The fifteen stories were selected from a pool of 32 stories initially constructed. The emotion content of the stories was assessed by twenty undergraduate psychology students who were asked to supply the emotion word or words that corresponded to how they thought the main character was most likely to be feeling given the scenario presented. The stories selected for the experiment were those for which the target emotion or a matching synonym was the most frequently supplied emotion-state.

Stories were designed to be interesting and enjoyable for children capable of reading text at the 7.5-year level and above. Within each set, all stories were matched exactly on number of words and number of sentences. Reading ability required for reading of each story was assessed using the Hatcher book grading system (Hatcher, 2000). Stories in each set of three varied by no more than three months in difficulty assessed in terms of reading age (as assessed by Wechsler Objective Reading Dimensions Basic Reading) required to read them. Details of the gradings and reading

ages required for each story are shown in table 1. The stories are provided in full in Appendix 1.

*Table 2.1* Hatcher grading and *WORD* reading age equivalents for experimental stories.

Story Name	Emotion	Hatcher Grade	<i>WORD</i> reading age equivalent (years)
Bear!	Fear	17	7,0
Lost	Fear	17	7,0
Zoo Trip	Fear	17	7,0
Birthday Party	Sadness	17	7,0
Poor Mum	Sadness	19	7,3
No Dogs	Sadness	18	7,0
Art Competition	Happiness	17	7,0
Family History	Happiness	15	6,7
Netball Team	Happiness	15	6,7
French Food	Disgust	19	7,3
School Hamster	Disgust	17	7,0
Living Lettuce	Disgust	17	7,0
Pink Bedroom	Anger	18	7,0
Test Cheat	Anger	17	7,0
Shopping	Anger	19	7,3
Mean Book Grade			
for experimental stories (SD)		17.34 (1.37)	7,0

For each emotion there were three target sentences, one embedded in each story. One story contained a consistent target sentence where the emotion-state of the main character was coherent with the context. One story contained an inconsistent target sentence, which was identical to the consistent target sentence except for the emotion-state word which was the perceived opposite to what would be predicted

from the context. The third story contained a target sentence that made no reference to emotion, but varied from the consistent target sentence by no more than two words.

For example, for sadness the target sentences were as follows:

Consistent: *Ben was very sad because he could not go*

Inconsistent: *Ben was very glad because he could not go*

Neutral: *Ben was ringing them because he could not go*

In this example, for all three stories the consistent emotion-state of the protagonist was sadness.

It was decided to use syllable reading times as the dependent measure, in order that small differences between target sentences in the number of syllables could not be found to account for any differences in reading times between target sentences.

However, it is important to recognise that the total target sentence reading time reflects both time spent actually reading the sentences, and time spent on additional processing before advancing to the next sentence. Therefore, using syllable reading time as the dependent measure makes it less likely that significant differences in processing time due to inconsistency would be shown; as such it could be argued that syllable reading time is a more stringent measure. It is not possible using the current methodology to ascertain exactly how much time was allocated to different tasks, however, future studies combining this methodology with eye tracking or neuroimaging techniques would overcome this.

The stories were read aloud by the children, in order that reading errors could be corrected, and to ensure that all of the story content had been read. It was also noted that this was in keeping with typical guided reading activity that children of this age are asked to complete in school.

### 2.2.1.2 Comprehension Questions

To check that children were reading for meaning and to provide an off-line index of inference processing, four comprehension questions were asked after each story – two asked for factual information from the story, and two required participants to generate an inference to obtain the correct answer; one required a coherence inference and one an elaborative inference. Both coherence and elaborative inferences require participants to integrate background knowledge with factual information presented in the text, however in the case of coherence inference, the inference must be generated in order for the passage to be understood. A coherence inference can take the form of anaphora or pronoun resolution, or bridging inferences required to fill information gaps in the text. In the case of elaborative inference, although the passage can be understood without the inference, the mental representation of the text and the resulting situation model will be more detailed when the inference is made. Thus, elaborative inferences build on the richness of the situation model constructed as a passage is read.

An example of one passage used and the comprehension questions that followed it is provided below. A full list of the comprehension questions is provided in Appendix 2.

*Holly was on a school trip. Her class were going to the zoo as part of 'wild week'. They saw elephants, monkeys, zebras and tigers. Holly wanted to see the lions most of all. She could see them, but they were all asleep. "Wake up!" shouted Holly. Suddenly, one of the lions jumped up and ran towards her. The lion ran up to the fence where Holly stood and did a huge roar! Holly had never been so scared before. She ran away fast. "He didn't want to be woken up!" said her teacher.*

Who were asleep? (Question requiring literal information)

What did the lion do after Holly shouted? (Question requiring literal information)

Why did the lion roar at Holly? (Coherence inference required for full understanding of the passage)

What do you think “Wild Week” was about? (Elaborative inference for richer understanding but not required for full understanding of the passage)

The questions used were selected from a pool of questions constructed by the experimenter to assess comprehension of the stories. The question types were rated by a sample of ten expert raters from the Centre for Reading and Language at The University of York, all of whom had considerable expertise in conducting reading research. Each rater was asked to state whether each question asked for literal information, or whether an inference had to be generated in order to answer the question. If an inference was required, raters were asked whether it was a coherence or elaborative inference. A coherence inference was defined as “necessary for understanding the text presented” and it was explained that this could either be in the form of a bridging inference required for gap filling (for example, to fully understand the sentence “Michael’s mum changed his nappy and gave him a bottle of milk” you would need to infer that Michael was a baby) or an inference required based on linguistic cues in the text (for example to understand “Tom hit Janet with his tennis racket”, you would need to infer that ‘his’ referred to the tennis racket that he was holding). For a question to be included in the study, the question type had to be agreed upon by the experimenter and a minimum of 8 out of the 10 expert raters. A one tailed binomial test shows that  $p=0.05$  for the minimum level of agreement required from the raters alone, therefore for the purpose of this study this criterion was deemed to be acceptable. Reliability of the off-line assessment was found to be

acceptable with Cronbach's  $\alpha = 0.75$  for the comprehension questions administered.

The questions administered can be found in full in Appendix 1.

#### 2.2.1.3 *Standardised Tests of Reading and Language Ability*

Children were screened and assessed using a range of published standardised tests and measures. In selecting these tests, test reliability and validity were carefully considered, along with current common practice in United Kingdom reading research.

In line with most recently published studies in the United Kingdom, reading accuracy and reading comprehension were measured using the *Neale Analysis of Reading Ability – Second British Edition* (NARA-II; Neale, 1997). In this test, children are required to orally read up to six passages that get progressively more difficult, and to answer 4 or 8 usually open-ended comprehension questions after each passage which tap inferential and literal information processing. This test was standardized using a representative sample of children in the United Kingdom, and internal consistency reliability of both forms of the test were found to be good for reading accuracy (Cronbach's coefficient alpha  $>0.8$  for all age groups tested) and very good for reading comprehension (Cronbach's coefficient alpha  $>0.9$  for all age groups tested). This compares very favourably with other standardized measures of reading comprehension such as WORD Reading Comprehension, and Gray Oral Reading Test version 4. Adequate evidence for construct validity of this test is shown by examining age-related change in test performance, reaching ceiling at approximately 13 years of age, this being the upper limit of the age range measured by this test.

Nonword reading, sight word reading and word reading fluency were measured using the *Test of Word Reading Efficiency (TOWRE, Torgesen, Wagner, & Rashotte, 1999.)* This test consists of two subtests, yielding standard scores for each

(sight word reading and phonemic decoding efficiency), and a combined standard score (Word reading efficiency standard score). In *Sight word reading*, children are asked to read as many of the words presented on a card as they can in 45 seconds. The words are commonly occurring English words that very gradually increase in difficulty. In *Phonemic Decoding*, children are asked to read as many of the nonwords presented on a card as they can in 45 seconds. The nonwords very gradually increase in difficulty. This test was selected because it is a reliable and valid measure of the constructs under examination (reliability coefficients  $>0.90$ ), and it is widely used, very quick to administer, and most children seem to enjoy the timed element to the task.

Single word reading ability was measured using the *Basic Reading* subtest of Wechsler Objective Reading Dimensions (WORD; Wechsler, 1990) to calculate standard scores and reading age equivalents. In this test, children are required to read a series of English words that gradually increase in complexity. WORD *Basic Reading* is a reliable and valid measure of word reading accuracy, with reliability and validity estimates in line with other standardised single word reading tests such as the British Ability Scales *Word Reading* subtest. Additionally, this test was chosen because reading age equivalents from this test had been calibrated with the Hatcher gradings used to assess the difficulty of the stories in the experiment. Thus, it was possible to say that a child with a reading age equivalent of 7 years 6 months on WORD Basic Reading was likely to be able to read a story at Hatcher Book Level 20 or below.

This test was adapted for use for this experiment, with 13 additional emotion descriptors included to check that the children were able to read the emotion words contained in the stories. A word reading card was created with these emotion words interspersed with the test words. It was necessary to intersperse the emotion words

with other single words so that the children were not made aware that these words were pertinent to the experiment. Reading age equivalents and standard scores were calculated for all participants.

IQ was measured using the short form (2 subtests) of Wechsler Abbreviated Scales of Intelligence (WASI) comprising a verbal test of vocabulary and the non-verbal matrix reasoning test (Wechsler, 1999). This test was selected because it is quick to administer, but also provides a reliable and valid estimate of general cognitive ability. The various forms of this test (2- and 4-subtest full scale IQ, Verbal (2 subtest) IQ, Performance (2 subtest) IQ) are widely used by researchers for participant evaluation in this type of study.

### 2.2.2 Design

A mixed design was used with one between-subjects factor, Group (typically developing children, poor decoders, poor comprehenders) and two within-subjects factors, Consistency (neutral, consistent, inconsistent), and Emotion (happiness, sadness, fear, anger, disgust). Each child read the same fifteen stories divided over two testing sessions. As stories were presented in two sessions, each story was allocated to one of two sets. There were eight stories presented in the first and seven presented in the second set. The order of presentation of each set of stories was counterbalanced with the same number of participants reading set one and set two first. Stories were pseudorandomly allocated to each set, with the stipulation that consistent and inconsistent stories for a given emotion could not be included in the same set. Within each set, stories were presented in random order for each participant, managed by *e-prime*.

Within each story, participants were required to read a target sentence. Ten of the target sentences described the emotion-state of the protagonist, in half this was



consistent with context, in half it was inconsistent. In the remaining five sentences the target sentences contained no reference to the emotion state of the protagonist.

### *2.2.3 Participants*

202 children in Years 4 and 5 of two York primary schools were screened using the Neale Analysis of Reading Ability (NARA-II) – Second British Edition (Neale, 1997). From this initial screening, three chronological age-matched groups of twenty-five children were selected to take part in the study. Twenty-five poor comprehenders were identified who satisfied all of the following criteria; comprehension age at least nine months below chronological age, reading accuracy age no more than six months below chronological age and at least twelve months higher than comprehension age. These criteria were chosen to ensure that participants did have a demonstrable discrepancy between reading accuracy and reading comprehension, and that reading accuracy was unimpaired. Twenty-five poor decoders were selected. Children were assigned to this group if their reading accuracy age was at least 12 months below chronological age, their comprehension age was in line with or above their reading age, and their reading accuracy was no lower than 7.5 years (in order that they were capable of reading the stories). These criteria were intended to broadly mirror those for the selection of poor comprehenders. It was noted that reading comprehension in these children would be constrained by decoding skill, therefore comprehension age at or above reading accuracy age was required for selection to take part in the study. The very poorest readers tested could not be selected for the study, as they needed to be capable of reading the test passages. Twenty-five typically developing readers were selected with reading and comprehension ages within 6 months of chronological age.

Of the twenty-five typical readers selected, three did not complete the study, five of the poor comprehenders did not complete the study and nine of the poor decoders did not complete the study. The reasons for non-completion are summarised in table 2.

*Table 2.2. Reasons for participant withdrawals from study*

Reason for non-completion	Poor comprehenders	Typical readers	Poor decoders
Moved away from participating school	2		
Long-term illness	1	1	
Unavailable for testing		2	1
Behavioural issues	1		
Task too difficult			6
Unable to read emotion words			2
Unwilling to complete second part of study	1		

In the poor decoder group, the number of participant withdrawals was high as in many cases the reading age of the child was only just above the threshold level for eligibility. Also, WORD basic reading is not a timed test, it only provides an indication of whether a child successfully reads a word, and not how arduous they found the task. Six poor decoders were withdrawn from the study by the experimenter, either because they were making a large number of reading errors, or because their decoding was very slow and effortful, and the length of the task would have led to frustration and distress. The study sample comprised 20 poor comprehenders, 22 typical readers and 16 poor decoders. Characteristics of these children are shown in table 3.

Table 2.3. Characteristics of participants by group

Measure	Group		Poor Comprehenders (n=20)		Typically developing children (n=22)		Poor decoders (n=16)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Chronological Age (months)	112	6.6	110	7.9	113	7.3		
NARA(II) Reading Accuracy (Form 2) <sup>1</sup>	101	5.2	98	3.2	87	4.9		
NARA (II) Reading Comprehension (Form 2) <sup>1</sup>	87	4.6	97	3.9	86	7.84		
Short form (2 subtest) IQ (Wechsler Abbreviated Scale of Intelligence (WASI)) <sup>1</sup>	98	8.2	106	10	101	11.8		
Test of Word Reading Efficiency (TOWRE) Sight Word Reading <sup>1</sup>	110	9.8	105	8.2	96	7.5		
TOWRE Phonemic Decoding Efficiency <sup>1</sup>	115	12	104	10.1	93	5.5		
Wechsler Objective Reading Dimensions (WORD) Basic Reading <sup>1</sup>	101	8.3	101	8.3	87	6.8		

<sup>1</sup>Standard Scores

There were no significant between group differences in chronological age, WASI vocabulary, WASI Matrix Reasoning, or overall WASI short form IQ. There was a significant main effect of group in NARA-II reading accuracy ( $F(2,55)=42.22$ ,  $p<.001$ ,  $\omega^2=.59$ ), with post-hoc analysis showing that this was accounted for by impaired performance on these tests by the poor decoder group. There was a significant main effect of group on NARA-II reading comprehension ( $F(2,55)=23.261$ ,  $p<.001$ ,  $\omega^2=.43$ ), with post-hoc analysis indicating that this was accounted for by superior comprehension in the typical readers. As expected the poor

decoders' comprehension was equivalent to that of the poor comprehenders due to their impaired reading accuracy.

#### *2.2.4 Procedure*

Participants were seen twice, the first test session lasted approximately 45 minutes, and the second test session lasted approximately 30 minutes. Test sessions were no more than one month apart.

During the first testing session participants were administered the *Basic Reading* subtest of Wechsler Objective Reading Dimensions (WORD) and both subtests of the Test of Word Reading Efficiency (TOWRE). Ability to read thirteen emotion words (anger, angry, fear, frightened, scared, happy, glad, delighted sad, upset, disgusted, brave, calm) was checked by embedding them within the WORD Basic Reading test. Children who were able to read less than 10 of the emotion words were also excluded from further participation.

Participants were then asked to read aloud the first set of stories. The stories were presented sentence by sentence on a Dell Inspiron 5100 laptop computer with intrinsic mouse. Administration of the experiment and collection of reading time data were controlled using *e-prime* experimental software. Instructions were presented on the screen and read aloud by the experimenter. A self-paced reading paradigm was employed. Participants were required to press a mouse button to advance to the next sentence as they read. The computer was set up so that pressing either the left or the right mouse button advanced the story to the next sentence. Reading times were recorded as time between mouse button presses to the nearest millisecond, and for the purpose of data analysis these were transformed into milliseconds per word. At the end of each story a simple picture relevant to the story appeared on the screen, and the experimenter asked four questions to check participants' comprehension of the story.

After the first set of stories was completed, the short form (*Vocabulary* and *Matrix Reasoning* subtests) of Wechsler Abbreviated Scales of Intelligence (WASI) was administered. If there was insufficient time available, or participants were having difficulty maintaining concentration, one or both of the WASI subtests was administered in the second test session. In the second testing session, participants read the second set of stories in exactly the same way as outlined for the first testing session.

After each story, four comprehension questions were asked. The answers supplied by the children were scored either 1 (correct) or 0 (incorrect) by the experimenter. After reading all of the stories, participants' understanding of the target emotions was checked by giving a short multiple-choice test of facial emotion recognition. Ten pictures from the Ekman 60 faces test (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002) were selected, five female and five male, each depicting one of the five emotions, happiness, sadness, fear, anger and disgust. Children were required to select (from the five options happy, sad, scared, angry and disgusted) the facial expression of emotion portrayed. Participants were also asked to define the nine emotion words used in the experimental stories (happy, sad, scared, angry, disgusted, upset, glad, brave, delighted). No significant between- group differences in emotion understanding were observed in either modality ( $F < 1$ ).

## *2.3 Results*

### *2.3.1 Comprehension Questions*

Participants' responses to the questions asked at the end of each story are shown in table 3. Overall, the children performed well answering the questions, indicating that they were reading for meaning.

*Table 2.4. Means (and Standard Deviations) of scores to responses to questions asked at the end of each story.*

Question Type	Poor Comprehenders (N=20)	Typically Developing Children (N=22)	Poor decoders (N=16)
Factual (/30)	25.35 (2.58)	26.27 (2.37)	27.94 (1.48)
Inference (/30)	23.00 (3.74)	24.86 (2.01)	25.00 (2.48)

The comprehension question scores were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was question type (factual, inferential). The analysis revealed a significant main effect of question type ( $F(1,55)=42.57, p < .001, \eta^2=.436$ ). Children were better at answering questions that asked for literal information, than those requiring the generation of an inference. There was also a significant main effect of group ( $F(1,55)=4.99, p < .05, \eta^2=.153$ ). Post-hoc analysis using the Tukey-Kramer approach revealed that the poor decoders were significantly better than the poor comprehenders at answering the comprehension questions. The typically developing children performed at a similar level statistically to the poor decoders and the poor comprehenders. The Group x Question Type Interaction did not reach significance ( $F(2,55)=1.71, p > .05$ ).

Half of the inference questions could be answered correctly if the appropriate coherence inference was generated, and the other half required generation of an elaborative inference. The responses to the inference questions were analysed by inference types as follows.

*Table 2.5. Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story.*

Question Type	Poor Comprehenders (N=20)	Typically Developing Children (N=22)	Poor decoders (N=16)
Cohesive (/15)	12.10 (1.92)	13.00 (1.35)	12.81 (1.60)
Elaborative (/15)	10.90 (2.55)	11.91 (1.80)	12.19 (1.42)

Responses to the inference questions were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was inference question type (coherence, elaborative). The analysis revealed a significant main effect of question type ( $F(1,55)=10.252, p < .01, \eta^2=.157$ ). Children were better at answering questions that required a coherence inference, than those requiring the generation of an elaborative inference. There was a trend for this to be less pronounced in the poor decoders, however, the main effect of group ( $F(2,55)=2.83, p < .1, \eta^2=.09$ ) and the interaction between group and inference type did not reach significance ( $F < 1$ ).

### *2.3.2 Reading times to target sentences*

Mean syllable reading times and standard deviations are shown in figure 2.1.

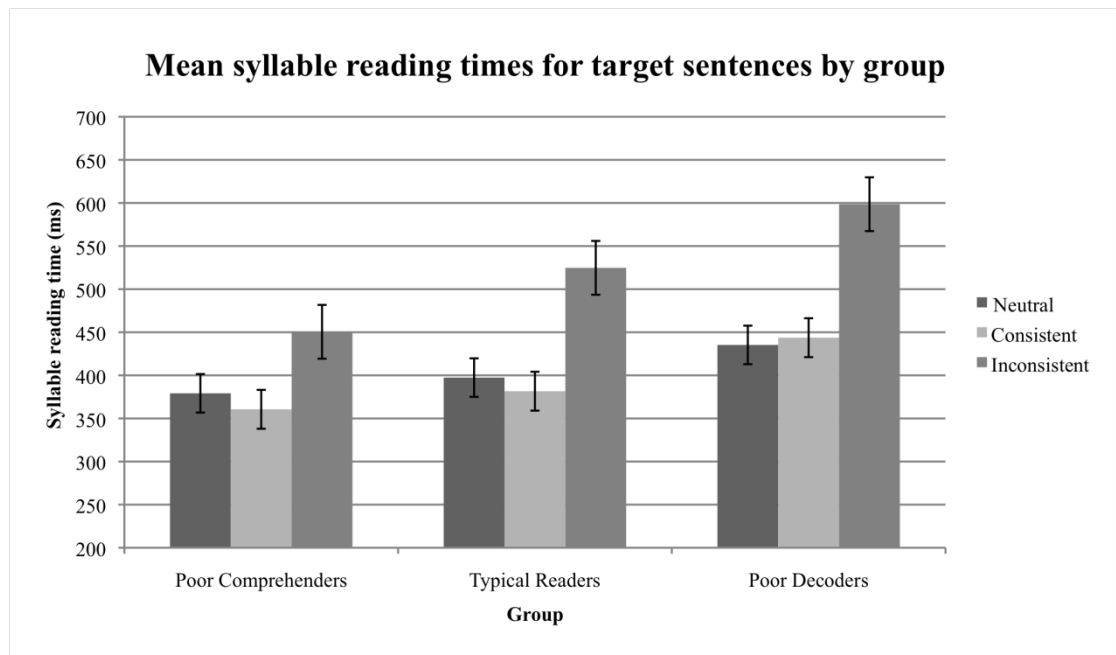


Figure 2.1. Mean syllable reading speeds to target sentences in milliseconds (\* error bars denote 95% confidence intervals)

Mean syllable reading times were entered into a 3 x 3 Mixed ANOVA. The between-subjects variable was group, which had three levels (poor comprehenders, typically developing children, poor decoders), and the within-subjects variable was consistency, which had three levels (neutral, consistent, inconsistent). Results are reported by subjects only, and not by items. Raaijmakers, Schrijnemakers and Gremmen (1999) provide validation for analyzing the data by subjects only, showing that by items analyses are not indicated where the items are well-matched on a range of variables. Throughout this thesis, analyses will be presented by subjects only.

The ANOVA revealed a highly significant main effect of Consistency ( $F(2,110) = 81.645, p < .001, \eta^2 = .597$ ). There was also a significant main effect of Group ( $F(2,55) = 5.858, p < .01, \eta^2 = .176$ ) and a significant Consistency x Group Interaction ( $F(4,110) = 3.004, p < .05, \eta^2 = .098$ ).

Post-hoc analysis using the Tukey-Kramer approach showed that reading of target sentences was inhibited in all groups when an inconsistency was present, and



that poor decoders were significantly slower at reading the target sentences than the poor comprehenders and marginally slower than the typical readers ( $p < .1$ ). Post-hoc analysis of the interaction between consistency and group showed that reading in the poor comprehenders was less disrupted by inconsistency than that of the typical readers or the poor decoders. In order to investigate this effect in the poor comprehenders further, a 2 x 2 ANOVA was conducted comparing the poor comprehenders and typical readers on the reading of consistent and inconsistent target sentences because their reading times for consistent sentences were in the same range. The analysis confirmed a significant main effect of consistency ( $F(1,40)=82.57, p < .001, \eta^2=.674$ ) and a significant consistency by group interaction ( $F(1,40)=4.301, p < .05, \eta^2=.097$ ) but no main effect of group ( $F(1,40)=3.367, p > .05$ ). The analysis confirmed that the poor comprehenders showed less disruption than typical readers when they read a sentence that was inconsistent with the context of the narrative.

As a further control for the possibility that the reduced disruption observed in the poor comprehenders in the inconsistent condition was an artifact of their faster overall reading speed, an index of relative inhibition in syllable reading times was calculated for each participant by dividing the difference between inconsistent and neutral syllable reading speeds by neutral syllable reading speeds for each participant. Means and standard deviations for the inhibition scores are shown in table 2.6.

Table 2.6. Comparison of inhibition in syllable reading times by group

	Poor Comprehenders (N=20)	Typically Developing Children (N=22)	Poor decoders (N=16)
Mean inhibition (ms per syllable)	0.20	0.34	0.40
Standard deviation	0.22	0.19	0.29

The data were entered into a one way between-subjects analysis of variance. The analysis revealed a significant main effect of group ( $F(2,55)=3.41$ ,  $p < .05$ ,  $\omega^2=.09$ ). Post-hoc analysis using independent-samples t-tests indicated that the poor comprehenders showed significantly less inhibition of reading than the typically developing readers ( $t(40)=2.098$ ,  $p < .05$ ) and the poor decoders ( $t(34)=2.263$ ,  $p < .05$ ).

#### 2.4 Discussion

Participants read inconsistent target sentences more slowly than consistent or neutral target sentences, confirming the hypothesis that an inconsistency in the emotion-state presented in the text would lead to inhibition of reading. No difference was found between reading times to consistent or neutral sentences, therefore the hypothesis that prior activation of emotion state words would result in facilitation in the consistent condition can be rejected. Poor decoders were slower at reading the target sentences overall, however the results showed that their reading was disrupted by the presence of inconsistency to the same level as the typical readers, indicating that when decoding is successfully achieved, processing of emotion-state inferences is intact.

Poor comprehenders read all sentences slightly faster than typical readers, however this difference was not statistically significant. Their reading was less disrupted by emotion-state inconsistency, suggesting that their automatic processing

of emotion-state inferences is not as robust as in typically developing readers or poor decoders. Thus, these results support the hypothesis that poor comprehenders show impaired inferential processing whereas poor decoders' inferential processing is not affected by their decoding difficulties. It is possible that during the additional time spent on decoding, inference generation proceeds automatically leading to either a greater number of inferences being generated, or a more elaborate situation model being constructed. Consistent with this, poor decoders were marginally better at answering questions requiring elaborative inferences to be generated

The hypothesis that facilitation would be observed to consistent target sentences due to prior activation of emotion-state material was not supported by the results. This suggests that inference processing may proceed independently of decoding, and perhaps emotion-state inferences are not generated at the level of the individual emotion words in line with the findings of Gygax, Oakhill and Garnham (2003). Another reason why facilitation was not observed may be due to the difficulty levels of the passages. The stories needed to be accessible to the poor decoder group in order that their inferential processing could be measured. However, this means that the texts were very easy for the poor comprehender group and the typically developing group to read, and were below the level of difficulty that these children would be reading at school. Thus, the children were likely to have been reading the texts quickly for them to begin with, and the possibility of facilitation, and the size of any effect may have been limited by this. A further study varying the difficulty level of text would clarify this.

The children's performance on the comprehension questions lends further support to the between-group differences observed in reading times. Overall, the children performed well on the comprehension questions, however the poor

comprehenders were worse at correctly responding to the questions than the poor decoders. The differences between the poor comprehenders and the typical readers, and between the poor decoders and the typical readers were not significant. In general, the children were better at correctly answering literal questions. When answering inferential questions, the children found it easier to answer questions requiring a coherence inference than questions requiring an elaborative inference.

A limitation of the study reported here is that only one inference type was measured. It is possible that poor comprehenders may have difficulty making emotion-state inferences automatically, reflecting a difficulty in understanding emotion specifically. Experiment 2 was conducted using a different type of inference, in order to determine whether the results of Experiment 1 reflected a specific difficulty with emotion-state inferences, or whether poor comprehenders' generation of other inference types is also compromised.

### *2.5. Experiment 2*

As in experiment 1, this study was designed to address two main research questions; Firstly, during the reading of short narratives, do children aged 7-10 years infer information about size and distance based on their general knowledge of the world? It was hypothesized that, in line with the results of Experiment 1, inconsistent target sentences would be read more slowly than consistent sentences. This would support the view that a range of inferences are generated on-line and automatically during reading.

Secondly, how does reading difficulty affect generation of spatial inferences? In order to test this, typically developing readers were again compared with two reading disabled groups; poor decoders and poor comprehenders. In line with the

results of Experiment 1, it was hypothesized that poor comprehenders would exhibit less inhibition of reading when presented with an inconsistent sentence.

In this experiment inferences about size and distance were used. 51 of the original 59 participants that took part in Experiment 1 were successfully recruited to take part in this experiment.

## *2.6 Method*

### *2.6.1 Design.*

A mixed design was used with one between-subjects factor, Group (Typically developing children, poor decoders, poor comprehenders) and one within-subjects factor, Consistency (consistent, inconsistent). Each child read the same ten stories divided over two testing sessions. Within each story, participants were required to read a target sentence containing information that related to size of an object or distance, in the context of the story. In four sentences the spatial characteristic was consistent with the story context, and in four sentences the spatial characteristic was inconsistent with the story context. In the remaining two sentences there was no reference to any spatial characteristics.

### *2.6.2 Materials*

#### *2.6.2.1 Experimental Stories*

Four sets of two stories were constructed. All stories described situations that children of this age group might reasonably find themselves in. Towards the end of each story, there was a statement about a spatial characteristic that related to the context of the story. In four sentences the spatial characteristic was consistent with the story context, and in four sentences the spatial characteristic was inconsistent with the story context. In the remaining two sentences there was no reference to any spatial characteristics.

One story in each set contained a consistent target sentence where the spatial characteristic was coherent with the context. The other story in each set contained an inconsistent target sentence, which differed by one or two words from the consistent target sentence and was inconsistent with context. The inconsistency was in relation to size or distance. For example, in sets one and two, the target sentences were as follows:

Consistent: *Dad picked up the shirt and put it in the shopping basket.*

Inconsistent: *Dad picked up the sofa and put it in the shopping basket.*

Consistent: *Her next race was the eight hundred metres. In a few minutes Katie would be at the finish line.*

Inconsistent: *There were four miles to go. In a few seconds Katie would be at the finish line.*

The ten stories used were selected from a pool of 14 stories initially constructed. The stories were piloted in a sample of ten undergraduate psychology students using the same method as the experiment overall. The stories selected for the experiment were those where the effect of inconsistency was found to be greatest in the adult sample.

Stories were designed to be interesting and enjoyable for children capable of reading text at the 7.5-year level and above. Within each set, all stories were matched exactly on number of words and number of sentences. Reading ability required for reading of each story was assessed using the Hatcher book grading system (Hatcher, 2000). Stories in each set varied by no more than three months in difficulty from each other as assessed in terms of reading age (as assessed by Wechsler Objective Reading Dimensions Basic Reading) required to read them. Details of the gradings and reading ages required for each story are shown in table 2.7. The name of each story is

preceded by the set number. Only two of the four sets included a neutral story for this study, as it was demonstrated in Experiment 1 that there was no difference in reading time between the neutral and consistent conditions. However it was felt that this finding needed to be replicated before the neutral condition could be abandoned. The stories are provided in full in Appendix 2.

*Table 2.7.* Hatcher grading and *WORD* reading age equivalents for experimental stories.

Set and Story Name	Hatcher Grade	<i>WORD</i> reading age equivalent (years)
1. Cup Final	15	6,7
1. Surf's Up	13	6,6
1. Match Point	14	6,7
2. Sofa Shopping	18	7,0
2. Birthday Shopping	17	7,0
3. Chocolate Cake	15	6,7
3. Lost	14	6,7
4. Holiday fun	16	6,8
4. Fun run	16	6,8
4. Sports day	15	6,7
Mean Book Grade for experimental stories (SD)	15.26 (1.52)	7,0

#### *2.6.2.2 Comprehension Questions*

To check that children were reading for meaning and to provide an off-line index of inference processing, as in Experiment 1, four comprehension questions were asked after each story – two asked for factual information from the story, and two required participants to generate an inference to obtain the correct answer; one

required a coherence inference and one an elaborative inference. Those used were selected from a pool of questions constructed to assess comprehension of the stories using the same protocol as in experiment 1. The question types were rated by a sample of ten expert raters from the Centre for Reading and Language at The University of York. Raters were asked to state whether the questions asked for literal information, or whether an inference had to be generated in order to answer the question. If an inference was required, as before, the expert raters were asked whether it was a coherence or elaborative inference. As in experiment 1, the questions that were included in the experiment were those where agreement on question type was 80% or more i.e. 8 out of 10 raters were in agreement on the question type, and this question type was the same as that intended by the experimenter. The internal consistency of the resulting assessment of off-line comprehension was found to be very high with Cronbach's  $\alpha = 0.98$ . A list of the comprehension questions is provided in Appendix 2.

### *2.6.3 Participants*

51 of the 59 children that took part in Experiment 1 were available to take part in this experiment. Twenty were typical readers, 17 were poor comprehenders, and 14 were poor decoders, selected according to the protocol outlined in Experiment 1. Characteristics of the children taking part in this experiment are shown in table 2.8.



Table 2.8. Characteristics of participants by group

Measure	Group		Poor		Typically		Poor decoders	
			Comprehenders		developing		(n=14)	
			(n=17)		children (n=20)			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Chronological Age (months)	111.12	6.98	110.2	8.186	112.86	7.624		
NARA(II) Reading Accuracy (Form 2) <sup>1</sup>	100.47	4.39	97.50	3.17	87.86	4.70		
NARA (II) Reading Comprehension (Form 2) <sup>1</sup>	87.76	4.50	96.80	4.03	87.07	7.23		
Short form (2 subtest) IQ (Wechsler Abbreviated Scale of Intelligence (WASI)) <sup>1</sup>	98.06	8.96	106.60	9.83	102.43	10.80		
Test of Word Reading Efficiency (TOWRE) Sight Word Reading <sup>1</sup>	109.18	8.86	103.15	6.03	96.5	7.61		
TOWRE Phonemic Decoding Efficiency <sup>1</sup>	115.53	10.38	101.50	6.62	93.789	5.54		
Wechsler Objective Reading Dimensions (WORD) Basic Reading <sup>1</sup>	102.53	7.99	99.80	8.11	88.07	6.57		

<sup>1</sup>Standard Scores

A series of one-way analyses of variance were performed to check the matching of groups on chronological age, NARA-II reading accuracy and reading comprehension and IQ. There were no significant between group differences in chronological age, WASI vocabulary or WASI Matrix Reasoning. However, there were significant between group differences in IQ when the two components were combined ( $F(2,48)=3.47, p=.039, \omega^2=.09$ ) and post-hoc testing revealed that the poor comprehenders IQ was lower than that of the typically developing readers. There was no significant difference between the IQ of the poor decoders and the poor

comprehenders. There was a significant main effect of group in NARA-II reading accuracy ( $F(2,48)=40.01, p <.001, \omega^2=.60$ ), with post-hoc analysis showing that this was accounted for by impaired performance on these tests by the poor decoder group. There was a significant main effect of NARA-II reading comprehension ( $F(2,48)=19.46, p <.001, \omega^2=.42$ ), with post-hoc analysis indicating that this was accounted for by superior comprehension in the typical readers. As expected the poor decoders' comprehension was equivalent to that of the poor comprehenders due to their impaired reading accuracy. The results indicate that selection and matching of groups was adequate. WORD Basic Reading was administered prior to exposure to the experimental stories, to ensure that the children were capable of reading them, and these results provided further evidence that the groups were well matched. A one way Analysis of Variance revealed a significant main effect of group ( $F(2,48)=15.097, p <.001, \omega^2=.38$ ), with post hoc analysis indicating that the poor decoders obtained lower scores than either the poor comprehenders or the typical readers. Of interest were the results of the TOWRE subtests. One way Analyses of Variance revealed significant main effects of group in both the Sight Word Reading ( $F(2,48)=11.538, p <.001, \omega^2=.28$ ) and the Phonemic Decoding Efficiency ( $F(2,48)=22.576, p <.001, \omega^2=.54$ ). Post-hoc analysis using Tukey's HSD revealed that, in line with their performance on NARA Reading Accuracy, the poor decoders were worse than both the typical readers and the poor comprehenders on both TOWRE subtests. Interestingly, despite being well matched on NARA Reading Accuracy, the poor comprehenders scored more highly on these tests than the typical readers, indicating that they could decode nonwords and identify sight words faster than typical readers.

#### *2.6.4 Procedure*

Participants read all ten stories in a single sitting that lasted approximately 25 minutes. Participants were asked to read the stories aloud. The stories were presented sentence by sentence on a Dell Inspiron 5100 laptop computer with intrinsic mouse. Administration of the experiment and collection of reading time data were controlled using E-prime experiment management software. Instructions were presented on the screen and read aloud by the experimenter. A self-paced reading paradigm was employed. Participants were required to press a mouse button to advance to the next sentence as they read. The computer was set up so that pressing either the left or the right mouse button advanced the story to the next sentence. Reading times were recorded as time between mouse button presses to the nearest millisecond, and for the purpose of data analysis these were transformed into milliseconds per word. At the end of each story a simple picture relevant to the story appeared on the screen, and the experimenter asked four questions to check participants' comprehension of the story.

### *2.7 Results*

#### *2.7.1 Comprehension Questions*

Participants' responses to the questions asked at the end of each story are shown in table 2.9. Overall, the children performed well answering the questions, indicating that they were reading for meaning.

*Table 2.9. Means (and Standard Deviations) of scores to responses to questions asked at the end of each story.*

Question Type	Poor Comprehenders (N=17)	Typically Developing Children (N=20)	Poor decoders (N=14)
Factual (/20)	13.00 (1.41)	14.80 (2.17)	14.79 (2.05)
Inference (/20)	10.18 (2.13)	12.95 (1.96)	12.00 (2.77)

The comprehension question scores were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was question type (factual, inferential). The analysis revealed a significant main effect of question type ( $F(1,48)=59.895, p < .001, \eta^2=.555$ ). Children were better at answering questions that asked for literal information, than those requiring the generation of an inference. There was also a significant main effect of group ( $F(2,48)=8.289, p < .001, \eta^2=.257$ ). Post-hoc analysis using the Tukey-Kramer approach revealed that the poor comprehenders were significantly worse than the poor decoders or the typically developing children at answering the comprehension questions. The Group x Question Type Interaction did not reach significance ( $F(2,48)=1.08, p=0.35$ ).

Half of the inference questions could be answered correctly if the appropriate coherence inference was generated, and the other half required generation of an elaborative inference. The responses to the inference questions were analysed by inference type and are shown in table 2.10 as follows.

*Table 2.10.* Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story.

Question Type	Poor Comprehenders (N=17)	Typically Developing Children (N=20)	Poor decoders (N=14)
Cohesive (/10)	5.00 (1.32)	6.35 (1.23)	6.21 (1.53)
Elaborative (/10)	5.18 (1.74)	6.6 (1.47)	5.79 (1.67)

Responses to the inference questions were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was inference type (coherence, elaborative). The analysis revealed a significant main effect of group ( $F(2,48)=9.853$ ,  $p<.001$ ,  $\eta^2=0.291$ ) and post-hoc analysis using Tukey's HSD revealed that the poor comprehenders were significantly worse at answering the inferential questions than the typical readers. The main effect of inference type and the interaction between question type and group did not reach significance (in both cases  $F<1$ ).

### *2.7.2 Reading Times to Target Sentences*

Mean syllable reading times are shown in figure 2.2.

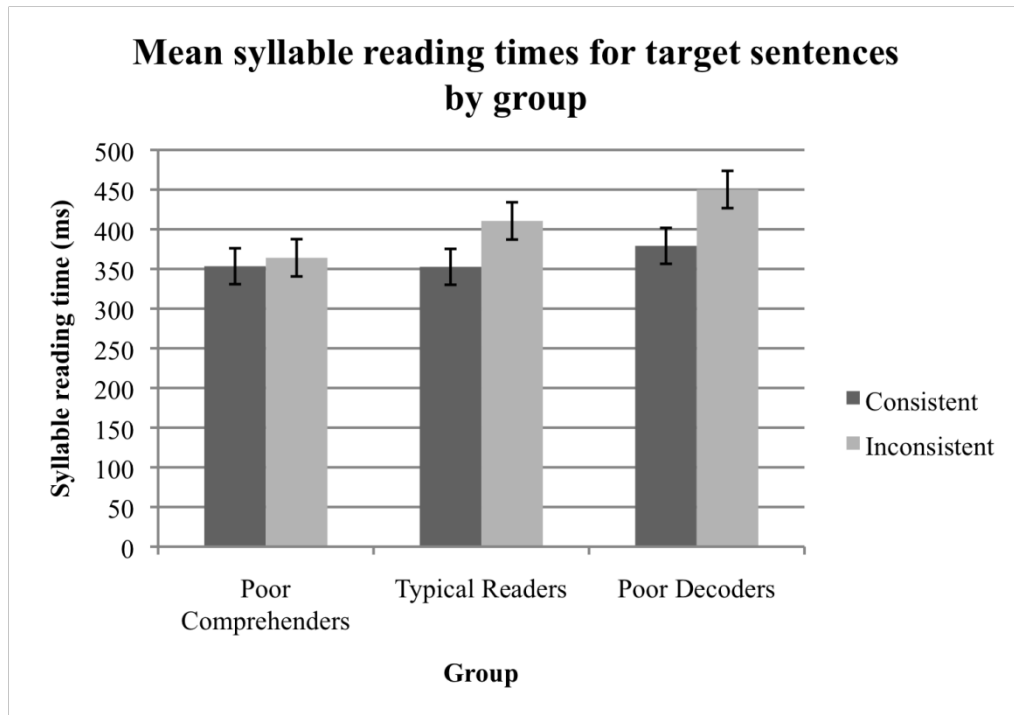


Figure 2.2. Mean syllable reading speeds to target sentences in milliseconds (\* error bars denote 95% confidence intervals)

Mean syllable reading times were entered into a 3 x 2 Mixed ANOVA. The between-subjects variable was group, which had three levels (poor comprehenders, typically developing children, poor decoders), and the within-subjects variable was consistency, which had two levels (consistent, inconsistent).

The ANOVA revealed a highly significant main effect of Consistency. ( $F(1,48) = 30.891, p < .001, \eta^2 = .392$ ) There was also a significant Consistency x Group Interaction ( $F(2,48) = 4.702, p < .05, \eta^2 = .164$ ). The main effect of group did not reach significance ( $F(2,48) = 2.085, p = .13$ ). Post-hoc analysis using the Tukey-Kramer approach showed that reading of target sentences was inhibited in all groups when an inconsistency was present. Post-hoc analysis of the interaction between

consistency and group showed that reading in the poor comprehenders was less disrupted by inconsistency than that of the typical readers or the poor decoders.

In order to investigate this effect in the poor comprehenders further, a 2 x 2 ANOVA was conducted comparing the poor comprehenders and typical readers on the reading of consistent and inconsistent target sentences. The analysis revealed a significant main effect of consistency ( $F(1,35)=11.723, p <.01, \eta^2=.251$ ) and a significant consistency by group interaction ( $F(1,35)=5.584, p <.05, \eta^2=.138$ ). The main effect of group was not significant ( $F(1,35)=1.199, p >.05$ ). The analysis confirmed that the poor comprehenders showed less disruption than typical readers when they read a sentence that was inconsistent with the context of the narrative.

In order to control for individual differences in reading speed, and thus exclude the possibility that the reduced disruption observed in the poor comprehenders in the inconsistent condition was due to them reading faster overall, an index of relative inhibition in syllable reading times was calculated for each participant. Means and standard deviations for the inhibition scores are shown in Table 2.11 below.

*Table 2.11. Comparison of inhibition in syllable reading times by group*

	Poor Comprehenders (N=17)	Typically Developing Children (N=20)	Poor decoders (N=14)
Mean inhibition (ms per syllable)	0.05	0.22	0.22
Standard deviation	0.113	0.236	0.170

The data were entered into a one way between-subjects analysis of variance. The analysis revealed a significant main effect of group ( $F(2,48)=4.654, p=.014, \omega^2=.13$ ). Post-hoc analysis using Tukey's HSD indicated that the poor comprehenders

showed significantly less inhibition of reading than either the poor decoders or the typically developing children.

## *2.8 Discussion*

Participants read inconsistent target sentences more slowly than consistent target sentences, confirming the hypothesis that an inconsistency in the spatial feature presented in the text would lead to inhibition of reading.

Poor comprehenders read all sentences slightly faster than typical readers, however this difference was not statistically significant. Their reading was less disrupted by the inconsistency, suggesting that their automatic processing of spatial inferences is not as robust as in typically developing readers or poor decoders. Thus, these results support the hypothesis that poor comprehenders show impaired inferential processing whereas poor decoders' inferential processing is not affected by their decoding difficulties.

The children's performance on the comprehension questions lends further support to the between-group differences observed in reading times. Overall, the children performed well on the comprehension questions, however the poor comprehenders were worse at correctly responding to the questions than the poor decoders. The differences between the poor comprehenders and the typical readers, and between the poor decoders and the typical readers were not significant. In general, the children were better at correctly answering literal questions. When answering inferential questions, the children found it easier to answer questions requiring a coherence inference than questions requiring an elaborative inference.



## *2.9 Summary and Conclusion*

Overall, the results from both experiments are in line with predictions based upon the Simple View of Reading proposed by Gough and Tunmer (1986). The differences between the groups are compatible with existence of dissociation between decoding and linguistic comprehension. In these experiments processing of both emotion-state inferences and spatial inferences was found to be intact in poor decoders. Poor comprehenders were less affected by the presence of an inconsistency in text suggesting that their inferential processing is less robust, or they attach less significance to inconsistencies in text. They clearly engage in less reanalysis when they encounter inconsistency, in line with studies showing reduced levels of comprehension monitoring in poor comprehenders (e.g. Cain, Oakhill & Bryant, 2004; Zinar, 2000). The reasons for this remain unclear. It is plausible that poor comprehenders become accustomed to texts not always making sense, and therefore they adopt a lower standard for coherence when monitoring their comprehension.

Conversely, these results are compatible with the possibility that poor decoders compensate for their decoding problems by attaching greater importance to comprehension processes such as inference generation to help improve their reading comprehension.

Taken together, the two studies demonstrate for the first time that typically developing children, poor decoders and to a lesser degree poor comprehenders, can and do make inferences about fictional characters' emotion-states and spatial inferences automatically during reading, providing they have sufficient skills to decode the text. Even in the presence of reading disability affecting decoding, these inferences are generated on-line as the text is processed.

Consistent with off-line studies showing that poor comprehenders generate fewer inferences than typical readers when they read, in both studies poor comprehenders showed less inhibition of reading when an inconsistency was presented in the text, suggesting that they either spent less time engaged in inferential activity attempting to resolve the inconsistency, or they failed to detect the inconsistency due to a failure to generate the relevant inferences during reading.

These studies extend the work of Gernsbacher and colleagues by applying the paradigm to developing readers. Furthermore, in the studies presented by Gernsbacher and colleagues, reading to consistent and inconsistent target sentences was compared between-subjects (Gernsbacher, Goldsmith & Robertson, 1992; Gernsbacher & Robertson, 1992; Gernsbacher, Hallada & Robertson, 1998). In these studies, the same effects were observed within-subjects, and these effects were maintained when individual differences in reading speed were controlled. This means that it is possible to use this technique to compare performance of typical and reading-disabled groups where reading speeds will vary, and to uncover factors that may constrain inferential processing.

A limitation in both experiments reported here was the difficulty level of the text. It is possible that the simplicity of the materials may have masked some of the poor comprehenders' difficulty with inference generation, and this may be better demonstrated using texts closer to their level of reading ability. Given that both studies provide converging evidence that poor decoders inferential processing is intact, it would be useful to conduct experiments using more difficult texts.

In conclusion, these studies demonstrate that like adults, children make inferences about fictional characters' emotion-states during reading, and that children make spatial inferences during reading. Comparison between performance of poor

comprehenders, poor decoders and typically developing readers has shown that this paradigm has potential to uncover information about inferential processing in typical and reading disabled groups. Results suggest that in children with decoding problems inferential processing is intact, and providing that decoding is successfully accomplished, inferences are generated normally. Poor comprehenders show reduced reanalysis of inconsistencies presented in the narratives, as their reading is less disrupted in the presence of such inconsistency. Further experiments using this paradigm may enable more information to be gleaned about constraints that operate on inferential processing and whether such constraints have any differential effect on poor comprehenders.

## CHAPTER 3

### **Knowledge Integration and On-line Inference Generation**

#### *3.1 Introduction*

Experiments 1 and 2 showed that, during reading, children automatically generate inferences about fictional characters' emotion-states, and about spatial features of objects or locations described in the text. These experiments have also confirmed that poor comprehenders have difficulty with inference generation. The propensity to integrate background knowledge with incoming text during reading varies according to a number of different parameters such as individual differences in age, reading skill, reading goals, and knowledge of reading strategies (McNamara & O'Reilly, 2009; Rapp, van den Broek, McMaster, Kendeou & Espin, 2007), and features of the text itself such as level of cohesion and difficulty (Ozuru, Dempsey & McNamara, 2009).

In Experiments 3 and 4 the issue of background knowledge is explored. More specifically, these experiments aim to examine whether the age at which knowledge of the world is acquired and specifically the recency of background knowledge predicts whether it can be utilized for the generation of inferences on-line during reading.

On a general level, domain knowledge plays an important role in determining how successfully an individual can extract meaning from a text. However, the ability to use that domain knowledge relies upon it being activated in response to incoming information, whether that information is presented in written format or auditorily.

Bransford and Johnson (1972; Experiments II & III) presented an elegant insight into the importance of activation of relevant domain knowledge when listening to a passage. Participants were read an ambiguous text containing information including the following:

“The procedure is actually quite simple. First you arrange things into different groups. Of course, one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities that is the next step, otherwise you are pretty well set” (p722; Bransford & Johnson, 1972).

Participants were either provided with the topic of the passage (washing clothes) before the text was read aloud to them, straight after the text was read aloud or were not provided with the topic at any point. They were then asked to recall as much information as possible from the passage, and to rate their comprehension of the passage. Participants recalled twice as many key ideas from the passages and rated their comprehension twice as highly when they had been told the subject of the passage prior to hearing it. Furthermore, supplying the topic immediately after the passage had been heard, but before the recall and rating tasks, had little impact on either participants’ performance on the recall task or the rating of their comprehension of the passage. This study provides a clear demonstration of the importance of integrating prior knowledge for the comprehension of an auditorily presented passage, but perhaps more crucially, it also emphasizes that in order for this to happen, the comprehender needs to be aware that prior knowledge needs to be activated and indeed which prior knowledge to activate.

According to Cain and Oakhill (1999) poor comprehenders are worse than controls at generating inferences that require them to integrate background knowledge and make connections with the information presented. This has also been demonstrated when the absolute level of background knowledge has been well controlled.

Cain, Oakhill, Barnes and Bryant (2001) developed an artificial knowledge base of facts about an imaginary planet called Gan. Poor comprehenders and controls

were taught 12 facts about Gan and its inhabitants, and their comprehension of a six-episode story was assessed once they could perfectly recall a set of background knowledge. In this way they ensured that participants from both groups had the same level of background knowledge, and the same knowledge base. After each of the six episodes, the children were asked four questions about the preceding episode that required them to supply either factual information or make an inference about what they had just heard. Less skilled comprehenders generated significantly fewer inferences of all types than the skilled comprehenders. This was despite both groups having the prerequisite knowledge to generate the inferences, ruling out the possibility that the poor comprehenders' difficulties were due to poor general knowledge. Rather, it was noted that their difficulty seemed to manifest at an earlier stage of the inference generation process. When skilled comprehenders experienced difficulty, it was more usually a failure to integrate items from the knowledge base with the relevant textual premise, but poor comprehenders had difficulty to begin with in actually selecting the correct information on which an inference needed to be based.

Skilled comprehenders learned the knowledge base more quickly than the less skilled comprehenders at the start of the study, and their recall of the knowledge base was superior to that of poor comprehenders one week after the study. Analysis of covariance demonstrated that this superior knowledge in itself did not account for poor comprehenders' failure to generate as many inferences. However, it is possible that, despite demonstrating equivalent absolute levels of background knowledge, poor comprehenders were less able to access the information required for inference generation in response to the questions posed. It has been suggested that poor comprehenders form less integrated and embellished mental representations of text,

and that this may make it harder for them to retrieve information from semantic memory when it is required (e.g. Perfetti, Landi & Oakhill, 2007).

One factor that may affect accessibility of a piece of information in the knowledge base is how recently it has been acquired. It seems reasonable to predict that information that has been used and reinforced over a considerable period of time will be more accessible than more recently acquired information because it should be more integrated into a network of connections within semantic memory. Experiments 3 and 4 explore whether the timing of acquisition of background knowledge makes a difference to whether it is accessible for on-line inference generation. The passages used in these experiments were constructed to induce children to generate particular inferences by integrating their general knowledge with the text.

In Experiment 3 half of the inferences that children were induced to generate rely on access to information children were typically familiar with at school entry, and the other half depended on information that the children were taught in the first term of Year 4. It was hypothesized that children would find it more difficult to generate an inference on-line when more recently acquired information had to be manipulated, therefore the main effect of consistency would be reduced in the less familiar condition. It was also hypothesized that this effect would be more pronounced in poor comprehenders.

### *3.2 Experiment 3 Method*

In Experiment 3, a self-paced reading paradigm was employed. Participants read twelve 70-100 word long passages, six of which included a target sentence consistent with inferences that could reasonably be inferred from integrating general knowledge with the passage content, and six included target sentences containing information inconsistent with such inferences. Syllable reading times to consistent

and inconsistent sentences were compared in typical readers, poor comprehenders and poor decoders, to examine the effect of reading difficulty on the ability to automatically integrate general with passage content.

### *3.2.1 Design*

A mixed design was used with one between-subjects factor, Group (Typically developing children, poor decoders, poor comprehenders) and two within-subjects factors, Consistency (consistent, inconsistent) and Familiarity (early, late). Each child read the same twelve stories in a single testing session. Within each story, participants were required to read a target sentence that related to information that may be inferred from the detail given in the story. In six sentences the content of the sentence was consistent with the inferred information, and in the other six sentences the content was inconsistent. Three consistent and three inconsistent sentences related to information that most children would know on school entry. The remaining three consistent and inconsistent sentences related to information that the children were taught in the first two terms of Year 4.

### *3.2.2 Materials*

#### *3.2.2.1 Experimental Stories*

Six sets of two stories were constructed. All stories described situations in which children of this age group might reasonably find themselves. Towards the end of each story, there was a statement that related to information that might reasonably be inferred from the earlier context. In six stories the target sentence was consistent with what might be inferred from the story, and in six sentences the target sentence was inconsistent with what might be inferred from the story.

The two stories in each set were matched very closely on number of sentences, sentence length and passage difficulty. The target sentences within these stories were



exactly the same length, and differed by no more than three words. The stories were constructed so that for each set, the target sentences were placed at the same point in the passage, with one target sentence consistent with context, and the other inconsistent with context. In order for this matching to be achieved, and for the context to be relevant to the inference being tested, the subject matter of the two stories in each pair was necessarily similar, but not identical. The general knowledge required for generation of the target inference was very similar across the two stories in each set.

The inconsistency related to information given earlier in the story, and could only be identified if an inference had been generated about that information. In sets 1-3 the background knowledge required to generate the inference was information that the children could reasonably be expected to know on school entry. Below is an example of a story containing an inconsistent target sentence (shown in italics).

Jack and dad took their cat, Oscar, to see the vet. Once a year, Oscar needed some vaccinations. These were special injections that stopped him from getting nasty diseases. He was very cross when they tried to put him in the cat basket. He put his claws out and wriggled furiously. Once he was safely in the basket, they put him in the car. All the way to the vets, he meowed at the top of his voice. When they got there, they took Oscar into the vet's office. The vet examined him, and gave him an injection. *He did not bark when the vet stuck the needle in.* The vet said he would see Oscar again next year. Jack and dad took Oscar home. He did not make a sound on the way back. "I think he's glad that the visit to the vet is over," said Jack.

Stories were designed to be interesting and enjoyable for children capable of reading text at the 8.5-year level and above. Reading ability required for reading of each story was assessed using the Hatcher book grading system (Hatcher, 2000). Stories in each set of two varied by no more than three months in difficulty from each other assessed in terms of reading age (as assessed by Wechsler Objective Reading

Dimensions Basic Reading) required to read them. Details of the gradings and reading ages required for each story are shown in Table 3.1. The stories are provided in full in Appendix 3.

*Table 3.1* Hatcher grading and *WORD* reading age equivalents for experimental stories.

Story Name	Background knowledge required	Hatcher Grade	<i>WORD</i> reading age equivalent (years)
Vet trip	Early	25	8,2
Sore throat	Early	25	8,2
Climbing frame fall	Early	23	7,9
Riding fall	Early	25	8,2
Lazy bones	Early	20	7,4
Sleepover	Early	20	7,4
Magnetic Apple	Late	25	8,2
Car boot sale	Late	25	8,2
Science lesson	Late	21	7,5
Vegetable patch	Late	20	7,4
Fat hedgehog	Late	22	7,7
Sleepy tortoise	Late	23	7,9
Mean Book Grade for experimental stories (SD)		22.66 (2.17)	7,9

### 3.2.2.2 Comprehension Questions

To check that children were reading for meaning and to provide an off-line index of inference processing, four comprehension questions were asked after each story – two asked for factual information from the story, and two required participants to generate an inference to obtain the correct answer; one required a coherence inference and one an elaborative inference. Both coherence and elaborative inferences

require participants to integrate background knowledge with factual information presented in the text, however in the case of coherence inference, the inference must be generated in order for the passage to be understood. In the case of elaborative inference, although the passage can be understood without the inference, the mental representation of the text and the resulting situation model will be more elaborate. The questions used were selected from a pool of questions constructed to assess comprehension of the stories using the same protocol as in experiment 1. The question types were rated by a sample of ten expert raters from the Centre for Reading and Language at The University of York. Raters were asked to state whether the questions asked for literal information, or whether an inference had to be generated in order to answer the question. If an inference was required, as before, the expert raters were asked whether it was a coherence or elaborative inference. As in experiment 1, the questions that were included in the experiment were those where agreement on question type was 80% or more i.e. 8 out of 10 raters were in agreement on the question type, and this question type was the same as that intended by the experimenter. . The internal consistency reliability of the resulting assessment of off-line comprehension was found to be very high with Cronbach's  $\alpha = 0.99$ . A list of the comprehension questions is provided in Appendix 3.

### *3.2.2.3 Standardised Tests of Reading and Language Ability*

The children taking part in this study had also participated in Experiments 1 and 2 and were selected according to the same protocol. Reading Accuracy and Reading Comprehension were measured using the *Neale Analysis of Reading Ability – Second British Edition* (NARA-II; Neale, 1997). Single word reading ability was measured using the *Basic Reading* subtest of Wechsler Objective Reading Dimensions (WORD; Wechsler, 1990). Nonword reading, sight word reading and reading fluency

were assessed using the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner & Rashotte, 1999). Reading age equivalents and standard scores were calculated for all participants.

IQ was measured using the short form (2 subtests) of Wechsler Abbreviated Scales of Intelligence (WASI) comprising a test of vocabulary and matrix reasoning (Wechsler, 1999).

### *3.2.3 Participants*

The 51 children that took part in Experiment 2 were invited to take part in this study. One child from the poor decoder group was unavailable for testing due to a family holiday, and testing was stopped in one typical reader as he was finding the stories very difficult to read, and did not wish to continue after reading five stories. There were 19 typical readers, 17 poor comprehenders, and 13 poor decoders. Characteristics of the children taking part in this experiment are shown in Table 3.2.

Table 3.2 Characteristics of participants by group

Measure	Group	Poor Comprehenders (n=17)		Typically Developing Children (n=19)		Poor Decoders (n=13)	
		Mean	SD	Mean	SD	Mean	SD
Chronological Age (months)		111.12	6.98	110.84	7.88	112.15	7.45
NARA (II) Reading Accuracy (Form 2) <sup>1</sup>		100.47	4.39	97.74	3.07	87.38	4.54
NARA (II) Reading Comprehension (Form 2) <sup>1</sup>		87.76	4.50	97.05	3.98	85.92	6.06
Short form (2 subtest) IQ (Wechsler Abbreviated Scale of Intelligence (WASI)) <sup>1</sup>		98.06	8.96	106.89	10.01	101.54	10.69
Test of Word Reading Efficiency (TOWRE) Sight Word Reading <sup>1</sup>		109.18	8.86	103.47	6.01	96.62	7.911
TOWRE Phonemic Decoding Efficiency <sup>1</sup>		115.53	10.38	102.21	5.968	93.62	5.723
Wechsler Objective Reading Dimensions (WORD) Basic Reading <sup>1</sup>		102.53	7.99	99.58	8.27	87.23	6.00

<sup>1</sup>Standard Scores

A series of one-way analyses of variance were performed to check the matching of groups on chronological age, NARA-II reading accuracy and reading comprehension and IQ. There were no significant between group differences in chronological age, WASI vocabulary or WASI Matrix Reasoning. However, there were significant between group differences in IQ when the two components were combined ( $F(2,46)=3.68, p=.033, \omega^2=.10$ ) and post-hoc testing revealed that the poor comprehenders IQ was lower than that of the typical readers. There was no significant

difference between the IQ of the poor decoders and the poor comprehenders. There was a significant main effect of group in NARA-II reading accuracy ( $F(2,46)=43.20$ ,  $p < .001$ ,  $\omega^2=.63$ ), with post-hoc analysis showing that this was accounted for by impaired performance on these tests by the poor decoder group. There was a significant main effect of NARA-II reading comprehension ( $F(2,46)=26.45$ ,  $p < .001$ ,  $\omega^2=.51$ ), with post-hoc analysis indicating that this was accounted for by superior comprehension in the typical readers. As expected the poor decoders' comprehension was equivalent to that of the poor comprehenders due to their impaired reading accuracy. The results indicate that selection and matching of groups was adequate.

The WORD Basic Reading test was administered prior to exposure to the experimental stories, to ensure that the children were capable of reading them, and the results provided support for the selection criteria used for each group. A one way Analysis of Variance revealed a significant main effect of group ( $F(2,46)=16.129$ ,  $p < .001$ ,  $\omega^2=.36$ ), with post-hoc analysis indicating that the poor decoders obtained lower scores than either the poor comprehenders or the typical readers.

Of interest were the unexpected results of the TOWRE subtests. One way Analyses of Variance revealed significant main effects of group in both the Sight Word Reading ( $F(2,46)=10.06$ ,  $p < .001$ ,  $\omega^2=.27$ ) and the Phonemic Decoding Efficiency ( $F(2,46)=30.94$ ,  $p < .001$ ,  $\omega^2=.55$ ). Post-hoc analysis using Tukey's HSD revealed that the poor decoders were worse than both the typical readers and the poor comprehenders on both subtests, but, interestingly, the poor comprehenders scored more highly on the phonemic decoding test than the typical readers, indicating that they could decode nonwords and identify sight words faster than typical readers.

### *3.2.4 Procedure*

Participants read all twelve stories in a single sitting that lasted approximately 30 minutes. Participants were asked to read the stories out loud. The stories were presented sentence by sentence on a Dell Inspiron 5100 laptop computer with intrinsic mouse. Administration of the experiment and collection of reading time data was controlled using e-prime version 1.1 experiment management software. Instructions were presented on the screen and read aloud by the experimenter. A self-paced reading paradigm was employed. Participants were required to press a mouse button to advance to the next sentence as they read. The computer was set up so that pressing either the left or the right mouse button advanced the story to the next sentence. Reading times were recorded as time between mouse button presses to the nearest millisecond, and for the purpose of analysis these data were transformed into milliseconds per syllable. At the end of each story, a simple picture relevant to the story appeared on the screen, and the experimenter asked four questions to check participants' comprehension of the story.

## *3.3 Results*

### *3.3.1 Comprehension Questions*

Participants' responses to the questions asked at the end of each story are shown in Table 3.3. Overall, the children performed well answering the questions, indicating that they were reading for meaning.

*Table 3.3. Means (and Standard Deviations) of scores to responses to questions asked at the end of each story.*

Question Type	Poor Comprehenders (N=17)	Typically Developing Children (N=19)	Poor Decoders (N=13)
Factual (/24)	17.88 (2.71)	19.84 (1.83)	19.23 (2.20)
Inference (/24)	16.18 (2.88)	18.42 (1.74)	18.31 (2.29)

The scores for the comprehension questions were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was question type (factual, inferential). The analysis revealed a significant main effect of question type ( $F(1,46)=17.038, p < .001, \eta^2=.270$ ). Children were better at answering questions that asked for literal information, than those requiring the generation of an inference though the size of the difference was small. There was also a significant main effect of group ( $F(2,46)=5.417, p < .01, \eta^2=.191$ ). Post-hoc analysis using Tukey's HSD revealed that the poor comprehenders were marginally significantly worse than the poor decoders and significantly worse than the typically developing children at answering the comprehension questions. The Group x Question Type Interaction did not reach significance ( $F < 1$ ).

Half of the inference questions could be answered correctly if the appropriate coherence inference was generated, and the other half required generation of an elaborative inference. The responses to the inference questions were analyzed by inference type and the results are shown in Table 3.4 below.



*Table 3.4. Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story.*

Question Type	Poor Comprehenders (N=17)	Typically Developing Children (N=20)	Poor decoders (N=14)
Cohesive (/12)	5.00 (1.32)	6.35 (1.23)	6.21 (1.53)
Elaborative (/12)	5.18 (1.74)	6.6 (1.47)	5.79 (1.67)

Responses to the inference questions were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was inference type (coherence, elaborative). The analysis revealed a significant main effect of group ( $F(2,46)=4.949, p=.011, \eta^2=.177$ ) and post-hoc analysis using Tukey's HSD revealed that the poor comprehenders were significantly worse at answering the inferential questions than the typical readers. The main effect of inference type did not reach significance ( $F<1$ ). The interaction between question type and group was marginally significant, ( $F(2,46)=2.561, p=.088, \eta^2=0.1$ ), and of interest was the fact that the poor comprehenders were worse at the coherence inference questions, whereas the other groups had greater difficulty with the elaborative inference questions.

A further analysis was performed in order to establish whether there were any differences in performance on the comprehension questions that related to the familiarity of the information contained within the story. Participants' responses were entered into a 3 x 2 x 2 Mixed ANOVA. The between-subjects variable was group as

in previous analyses, and the within-subjects variables were question type (literal, inferential) and familiarity (early-acquired, late-acquired). The analysis revealed significant main effects of question type and group as identified previously in the analysis of literal and inferential question responses by group, and there was also a highly significant main effect of familiarity ( $F(2,46)=47.78$ ;  $p < .001$ ;  $\eta^2 = .509$ ). The group by question type ( $F < 1$ ), familiarity by group ( $F < 1$ ), question type by familiarity ( $F < 1$ ) and question type by familiarity by group ( $F < 1$ ) interactions were not significant. Post-hoc analysis revealed that participants were worse at answering comprehension questions in the stories that related to later-acquired information.

### 3.3.2 Reading Times to Target Sentences

Mean syllable reading times and standard deviations by condition and group are shown in Figure 3.1.

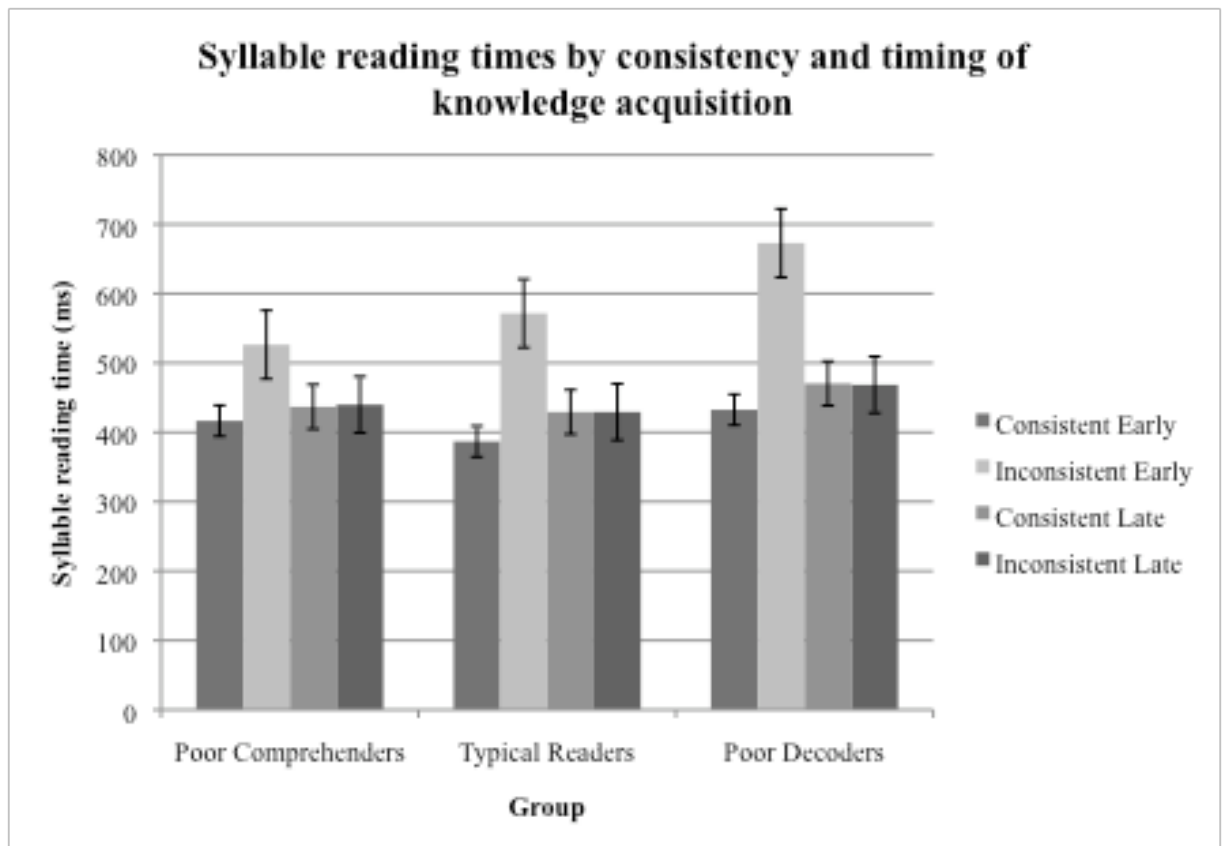


Figure 3.1. Mean syllable reading speeds to target sentences in milliseconds

(\* error bars denote 95% confidence intervals)

Mean syllable reading times were entered into a 3 x 2 x 2 Mixed ANOVA. The between-subjects variable was group, which had three levels (poor comprehenders, typically developing children, poor decoders), and the within-subjects variables were consistency, which had two levels (consistent, inconsistent), and familiarity (early-acquired, late-acquired).

The ANOVA revealed a highly significant main effect of Consistency ( $F(1,46) = 43.506, p < .001, \eta^2 = .486$ ). There was also a significant main effect of familiarity ( $F(1,46) = 19.892, p < .001, \eta^2 = .302$ ). The main effect of group was not significant ( $F(2,46) = 1.398, p = .26$ ).

The interaction between consistency and familiarity was highly significant ( $F(1,46) = 53.266, p < .001, \eta^2 = .537$ ). Neither the interaction between familiarity and group ( $F(2,46) = 1.27, p = .29$ ), consistency and group ( $F(2,46) = 1.71, p = .19$ ), nor the three-way interaction between consistency, familiarity and group ( $F(2,46) = 2.46, p = .10$ ) reached significance.

Post-hoc analysis of the interaction between consistency and familiarity showed that there was a large significant effect of consistency in the early-acquired conditions, and no significant effect of consistency in the late-acquired conditions. Reading times in the late acquired conditions were significantly slower overall.

In order to control for individual differences in reading speed, an index of relative inhibition in syllable reading times was calculated for each participant for the early- and late-acquired conditions. Means and standard deviations for the inhibition scores are shown in table 3.5.

Table 3.5. Comparison of inhibition in syllable reading times by group

	Poor Comprehenders (N=17)	Typically Developing Children (N=19)	Poor decoders (N=13)
Early acquired	0.312	0.551	0.575
Late acquired	0.038	0.020	0.020

The data were entered into a 3 x 2 mixed analysis of variance. The analysis revealed a significant main effect of familiarity ( $F(1,46)=44.546$ ,  $p < .001$ ;  $\eta^2 = .492$ ). In line with the reading time results, there was significantly more inhibition of reading in the presence of inconsistency in the stories that induced inferences about early-acquired information. There was no significant interaction between familiarity and group ( $F(2,46)=1.863$ ,  $p=.17$ ), and no significant main effect of group ( $F(2,46)=1.247$ ,  $p=.30$ ).

### 3.4 Discussion

Participants read inconsistent target sentences more slowly than consistent or neutral target sentences, confirming the hypothesis that a factual inconsistency presented in the text would lead to inhibition of reading. However this effect was limited to the early-acquired stories.

In this study, no between group differences were observed between the poor comprehenders, poor decoders and typical readers. This was contrary to expectation, as it was predicted that poor comprehenders would show a reduced effect of consistency, and that this would be more pronounced in the late-acquired condition. It is plausible that the early-acquired information was extremely easy for all groups to integrate with the situation model of the story so that the inconsistency was obvious for all, but that it was too difficult to integrate the late acquired information so

children did not notice the inconsistency there. Given that the knowledge acquisition time points compared were preschool and Year 4, it is suggested that materials containing knowledge acquired between these times (for example in years 2 and 3) might be more sensitive to between-group differences, in that poor comprehenders might find integration of knowledge acquired at these times more difficult than typical readers.

Performance on the comprehension questions supports the initial classification of the children; as expected, the poor comprehenders were worse at answering the questions than both the poor decoders and the typical readers who performed at a similar level.

The results obtained in this experiment suggest that inferences requiring integration of recently acquired general knowledge with passage context may not be generated automatically during reading. If the inferences were generated contemporaneously as the passage was read, one would predict that participants would spend longer reading sentences inconsistent with the prior context as they sought to resolve the inconsistency. Their failure to do so, suggests that they failed to notice the inconsistency because they did not generate the inference required for it to be detected. Given these participants' performance in Experiment 2, the alternative explanation, that they did notice the inconsistency but made no additional effort to resolve it, seems unlikely.

It is plausible, however that processing of inferences requiring the integration of recently acquired general knowledge takes longer, therefore effects on text reading times might be seen at a later stage in reading of the narrative. It is possible, therefore that evidence of on-line inference generation may be seen in the reading times to sentences occurring after the target sentence is read. In Experiment 3, the content of

the post-target sentences was not closely matched and in some cases the target sentence was the last or second to last sentence in the passage, therefore it was not possible to examine this possibility.

In summary, in Experiment 3, typical readers, poor comprehenders and poor decoders were found to generate knowledge-based inferences automatically on-line when the information to be integrated with passage context consisted of facts they had known for a number of years. However, contrary to prediction, this experiment did not find evidence that generation of knowledge-based inferences requiring integration of recently acquired general knowledge proceeds automatically; none of the groups showed increased reading times when inconsistencies were encountered in the text. Furthermore, the predicted between-group differences between the typical readers, poor comprehenders and poor decoders were not shown, therefore the hypothesis that poor comprehenders would have more difficulty with inferences requiring integration of later-acquired information was not supported.

### *3.5 Experiment 4*

#### *3.5.1 Introduction*

Experiment 4 was conducted to investigate the time course of on-line inference processing to determine whether predicted differences in reading times to consistent and inconsistent passages emerge later during text reading if late-acquired general knowledge needs to be integrated with the passage in order for an inconsistency to be processed. Experiment 4 was conducted to extend and replicate Experiment 3 with passages constructed to allow an analysis of post-target sentence reading times. In this experiment therefore, the same knowledge-based inferences were used as in Experiment 3, with half of the inferences reliant on information children should be familiar with at school entry, and the others reliant on information that the children

were taught in previous school years and specifically during term 1 of Year 4. It was hypothesized that all groups would find it more difficult to generate an inference on-line when recently acquired information had to be manipulated, thus replicating Experiment 3. It was also hypothesized that inferences requiring the integration of recently acquired knowledge would be generated more slowly by all groups, and therefore delayed consistency effects would be seen as evidenced from reading times to post target sentences.

### *3.6 Experiment 4 Method*

Extended versions of each of the twelve stories used in Experiment 3 were constructed. In each of the extended versions there was a minimum of three sentences to read after the target sentence containing the information that was consistent or inconsistent with the inferred information.

#### *3.6.1 Design*

A mixed design was used with one between-subjects factor, Group (Typically developing children, poor decoders, poor comprehenders) and two within-subjects factors, Consistency (consistent, inconsistent) and Familiarity (early, late). Each child read extended versions of the twelve stories used in Experiment 3 in two testing sessions. Within each story, participants were required to read a target sentence that related to information that may be inferred from the detail given in the story. As in Experiment 3, in six sentences the content of the sentence was consistent with the inferred information, and in the other six sentences the content was inconsistent. Three consistent and three inconsistent sentences related to information that the children would have been expected to have known on school entry. The remaining three pairs of consistent and inconsistent sentences related to information that the children were taught in previous school years and in the first term of Year 4. The

class teachers of participants in this study confirmed that the late-acquired content had been taught in the curriculum during the first term of the present school year.

Readings were taken of reading times to the target sentence containing the information requiring generation of an inference and to the three sentences that followed.

### 3.6.2 *Materials*

#### 3.6.2.1 *Experimental Stories*

Extended versions of the twelve stories used in Experiment 3 were constructed. The stories were exactly the same as those used in Experiment 3 up to and including the target sentence. The stories were extended so that after each target sentence there was a minimum of 3 post-target sentences and across each consistent/inconsistent story pair, each post-target sentence differed by no more than one or two words, so that syllable reading times could be compared.

Below is the extended version of the example story used previously. In this story the target sentence is inconsistent with the story content and relates to factual knowledge (cats make a ‘meow’ sound) that children would be expected to have acquired by school entry.

Jack and dad took their cat, Oscar, to see the vet. Once a year, Oscar needed some vaccinations. These were special injections that stopped him from getting nasty diseases. He was very cross when they tried to put him in the cat basket. He put his claws out and wriggled furiously. Once he was safely in the basket, they put him in the car. All the way to the vets, he meowed at the top of his voice. When they got there, they took Oscar into the vet’s office. The vet examined him, and gave him an injection. *He did not bark when the vet stuck the needle in. Jack thought Oscar was very brave. The vet said he wanted to see Oscar again next year. He told them that Oscar would need another injection then.* Jack and dad took Oscar home. He did not make a sound on the way back. “I think he’s glad that the visit to the vet is over,” said Jack.

Within each set, all stories were closely matched on number of words and exactly matched on number of sentences. Reading ability required for reading of each



story was assessed using the Hatcher book grading system (Hatcher, 2000). As before, the stories in each set varied by no more than three months in difficulty assessed in terms of reading age (as assessed by WORD Basic Reading) required for decoding them. Details of the gradings and reading ages required for each of the extended versions of each story are shown in Table 3.6. The stories are provided in full in Appendix 4.

*Table 3.6.* Hatcher grading and *WORD* reading age equivalents for extended experimental stories

Story Name	Background knowledge required	Consistency	Hatcher Grade	<i>WORD</i> reading age equivalent (years)
Vet trip	Early	Inconsistent	25	8,2
Sore throat	Early	Consistent	25	8,2
Climbing frame fall	Early	Inconsistent	22	7,7
Riding fall	Early	Consistent	22	7,7
Lazy bones	Early	Inconsistent	22	7,7
Sleepover	Early	Consistent	23	7,9
Magnetic Apple	Late	Inconsistent	22	7,7
Car boot sale	Late	Consistent	21	7,5
Science lesson	Late	Inconsistent	23	7,9
Vegetable patch	Late	Consistent	22	7,7
Fat hedgehog	Late	Inconsistent	23	7,9
Sleepy tortoise	Late	Consistent	24	8,0
Mean Book Grade for experimental stories (SD)			23.00 (1.00)	7,8 (0,2)

### *3.6.2.2 Comprehension Questions*

To check that children were reading for meaning and to provide an off-line index of inference processing, four comprehension questions were asked after each story. All but three of the questions were the same as those used in Experiment 3; three of the questions requiring factual information were modified as the content of the story after the target sentence had changed very slightly. A list of the comprehension questions is provided in Appendix 3, with the amendments to the three questions provided at the end of the Appendix.

### *3.6.2.3 Standardised Tests of Reading and Language Ability*

Screening and Selection of participants to take part in this study was carried out in collaboration with a large-scale intervention study that was being carried out in 21 York schools.

Participant selection was conducted in two phases. A group-administered screening procedure was used to identify children to be screened individually. The measures used for this purpose were Spelling (an adapted version of Wechsler Objective Reading Dimensions Spelling (Wechsler, 1990)), Non-verbal IQ (Ravens Standard Progressive Matrices (Raven, 1998)), and Listening Comprehension (adapted from the Neale Analysis of Reading Ability II (Neale, 1997), form 1). Spelling was used as a proxy for decoding as it has been shown that spelling and phonological decoding are highly correlated, and this could be administered as a group task. The listening comprehension task had been developed and administered previously by Durand, Hulme, Larkin & Snowling (2005) who found reliability to be acceptable (Cronbach's Alpha =0.72).

From the group screening, children were selected for further individual assessment if their results suggested that they would fit one of the three reading skill profiles below:

- 1) Children with age-appropriate phonological decoding skills, non-verbal IQ and comprehension in the average range (typical readers)
- 2) Children with age-appropriate phonological decoding skills, non-verbal IQ in the average range, and weak comprehension (poor comprehenders)
- 3) Children with weak phonological decoding skills (but sufficient decoding skills to enable them to decode the experimental stories), non-verbal IQ in the average range (poor decoders).

Children from these three groups were then assessed individually. A similar test battery was employed to that used for Experiments 1,2 and 3, in order that valid comparisons of similarities and differences between participant groups could be made if required. Furthermore, these tests are extensively used in the selection of participants in UK reading research, with most featuring in the majority of recently published work (Hulme & Snowling, 2009).

Nonword reading, sight word reading and reading fluency were assessed using the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner & Rashotte, 1999). This test was selected for these experiments because it is a reliable and valid measure of the constructs under examination (reliability coefficients  $>0.90$ ), and it is widely used, very quick to administer, and most children seem to enjoy the timed element to the task. Reading age equivalents and standard scores were calculated for all participants.

Verbal IQ was measured using Wechsler Abbreviated Scales of Intelligence (WASI) comprising a test of vocabulary and verbal similarities (Wechsler, 1999). These tests were selected to provide a reliable and valid estimate of verbal ability. The various forms of this test (2- and 4-subtest full scale IQ, Verbal (2 subtest) IQ, Performance (2 subtest) IQ) are widely used by researchers for participant evaluation in this type of study. Passage Reading Accuracy and Reading Comprehension were assessed using form 2 of the Neale Analysis of Reading Ability (NARA-II) – Second British Edition (Neale, 1997) in line with the majority of comprehension research conducted in the United Kingdom. Phonological Skill was tested using an 18-item Phoneme Deletion task (after McDougall, Hulme, Ellis & Monk, 1994). A phonological measure was included in the test battery to provide additional validation that the poor comprehenders were different from the poor decoders on this marker of poor reading. This also provides further evidence to demonstrate that poor comprehenders did not show any phonological difficulties in addition to a deficit in linguistic comprehension.

### *3.6.3 Participants*

Three groups of fifteen children took part in this study. Characteristics of the children taking part in this experiment are shown in table 3.7.

Table 3.7. Characteristics of participants by group

Measure \ Group	Poor Comprehenders (n=15)		Typically Developing Children (n=15)		Poor decoders (n=15)	
	Mean	SD	Mean	SD	Mean	SD
Chronological Age (months)	104.27	4.68	102.13	3.94	104.27	4.70
NARA (II) Reading Accuracy (Form 2) <sup>1</sup>	100.33	6.91	102.40	5.34	88.20	2.96
NARA (II) Reading Comprehension (Form 2) <sup>1</sup>	84.27	7.72	99.13	4.87	86.53	3.68
Verbal IQ (Wechsler Abbreviated Scale of Intelligence (WASI)) <sup>1</sup>	95.87	14.37	110.07	6.85	101.73	11.07
Vocabulary (WASI) <sup>2</sup>	42.07	12.55	52.20	8.75	46.27	9.22
Verbal Similarities (WASI) <sup>2</sup>	51.67	7.54	59.27	6.38	56.00	6.36
Nonverbal IQ (Ravens Progressive Matrices) <sup>1</sup>	101.30	8.28	99.44	9.74	96.42	11.01
Listening Comprehension (Adapted from NARA (II) Form 1) <sup>1</sup>	88.09	11.71	102.97	14.33	97.99	10.88
Test of Word Reading Efficiency (TOWRE) Sight Word Reading <sup>1</sup>	106.27	12.42	106.80	7.19	95.27	8.51
TOWRE Phonemic Decoding Efficiency <sup>1</sup>	104.67	12.78	107.47	8.43	92.27	5.54
Phoneme Deletion (McDougall Test of Phoneme Deletion) (/18)	13.27	3.173	14.33	2.19	11.33	2.80

<sup>1</sup>Standard Scores

<sup>2</sup>T-scores

A series of one-way analyses of variance were performed to confirm the matching of groups. There were no significant between group differences in chronological age or Raven's Progressive Matrices. However, there were significant between group differences in Verbal IQ ( $F(2,42)=6.09, p < .01, \omega^2=.18$ ) and post-hoc testing revealed that the poor comprehenders' Verbal IQ was lower than that of the typically developing readers. There was no significant difference between the IQ of the poor decoders and the poor comprehenders and no significant difference between the Verbal IQ of the poor decoders and the typically developing readers. The same pattern of results was observed across both subtests within the Verbal IQ.

There were significant between-group differences in NARA-II reading accuracy ( $F(2,42)=31.12, p < .001, \omega^2=.57$ ), with post-hoc analysis showing that this was accounted for by impaired performance on these tests by the poor decoder group. There was a significant difference between groups in NARA-II reading comprehension ( $F(2,42)=29.80, p < .001, \omega^2=.56$ ), with post-hoc analysis indicating that this was accounted for by superior comprehension in the typical readers. It was assumed that the poor decoders' comprehension was equivalent to that of the poor comprehenders owing to their impaired reading accuracy being an obstacle to comprehension. This assumption was supported by the results of the Listening Comprehension test. There were significant between-group differences in Listening Comprehension ( $F(2,42)=5.60, p < .05, \omega^2=.17$ ), with post-hoc analysis indicating that the poor comprehenders performed significantly worse than the typical readers and marginally significantly worse than the poor decoders ( $p = .08$ ). There was no significant difference between the performance of the typical readers and that of the poor decoders on this test.

Further evidence in support of the classification of participants is provided by results of the TOWRE and McDougall Test of Phoneme Deletion. There were significant between-group differences in both the Sight Word Reading ( $F(2,42)=6.85$ ,  $p < .01$ ,  $\omega^2=.21$ ) and the Phonemic Decoding Efficiency ( $F(2,42)=11.11$ ,  $p < .001$ ,  $\omega^2=.31$ ). Post-hoc analysis using Tukey's HSD revealed that the poor decoders were worse than both the typical readers and the poor comprehenders on both subtests. A similar pattern of results was observed in the McDougall Test of Phoneme Deletion, in which poor decoders obtained significantly lower scores than either poor comprehenders or typical readers ( $F(2,42)=4.59$ ,  $p < .05$ ,  $\omega^2=.14$ ).

#### 3.6.4 Procedure

Participants read the twelve stories divided across two testing sessions. Each of the two sessions lasted approximately twenty minutes. Participants were asked to read the stories aloud. The stories were presented sentence by sentence on a Dell Latitude 120L laptop computer with intrinsic mouse. Administration of the experiment and collection of reading time data were controlled using E-prime version 1.1 experiment management software. Instructions were presented on the screen and read aloud by the experimenter. A self-paced reading paradigm was employed. Participants were required to press a mouse button to advance to the next sentence as they read. The computer was set up so that pressing either the left or the right mouse button advanced the story to the next sentence. Reading times were recorded as time between mouse button presses to the nearest millisecond, and for the purpose of data analysis these were transformed into milliseconds per syllable. At the end of each story a simple picture relevant to the story appeared on the screen, and the experimenter asked four questions to check participants' comprehension of the story.

### 3.7 Results

#### 3.7.1 Comprehension Questions

Participants' responses to the questions asked at the end of each story are shown in Table 3.8. Overall, the children performed well answering the questions, indicating that they were reading for meaning.

*Table 3.8.* Means (and Standard Deviations) of scores to responses to questions asked at the end of each story.

Question Type	Poor Comprehenders (N=15)	Typically Developing Children (N=15)	Poor decoders (N=15)
Factual (/24)	17.93 (2.43)	20.80 (2.27)	18.40 (5.58)
Inference (/24)	14.13 (2.92)	16.13 (2.59)	14.93 (4.94)

The comprehension question scores were entered into a three way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variables were question type (factual, inferential) and age of knowledge acquisition (early-acquired, late-acquired). The analysis revealed a highly significant main effect of question type ( $F(1,42)=99.86, p<.001, \eta^2=.70$ ). Children were better at answering questions that asked for literal information, than those requiring the generation of an inference. There was also a significant main effect of inference familiarity ( $F(1,42)=38.025, p<.001, \eta^2=.48$ ). Children were better at answering questions of both types when they referred to passages containing early-acquired inferences; poor comprehenders answered fewer questions correctly than either the poor decoders or typically



developing children overall, however the difference between their performance and that of the other groups was not significant. The main effect of group was not significant ( $F(2,42)=2.02$ ,  $p=.15$ ), and neither were the interactions between group and question type ( $F<1$ ), group and familiarity ( $F<1$ ), question type and familiarity ( $F(2,42)=1.28$ ,  $p=.26$ ), and group and familiarity and question type ( $F(2,42)=2.017$ ,  $p=.15$ ).

Half of the inference questions could be answered correctly if the appropriate coherence inference was generated, and the other half required generation of an elaborative inference. The responses to the inference questions were analyzed by inference types and are shown in Table 3.9.

*Table 3.9.* Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story.

Question Type	Poor Comprehenders (N=15)	Typically Developing Children (N=15)	Poor decoders (N=15)
Cohesive (/12)	6.80 (2.08)	7.80 (1.42)	7.33 (2.38)
Elaborative (/12)	7.33 (1.72)	8.33 (1.84)	7.60 (2.69)

Responses to the inference questions were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was inference type (coherence, elaborative). The main effects of group ( $F(2,42)=1.14$ ,  $p=.33$ ), and question type ( $F(1,42)=2.299$ ,  $p=.14$ ) and the interaction between group and question type ( $F<1$ ) were not significant.

### 3.7.2 Reading Times to Target Sentences

#### 3.7.2.1 Syllable Reading Times

Mean syllable reading times and standard deviations by condition and group are shown in Figure 3.2 below.

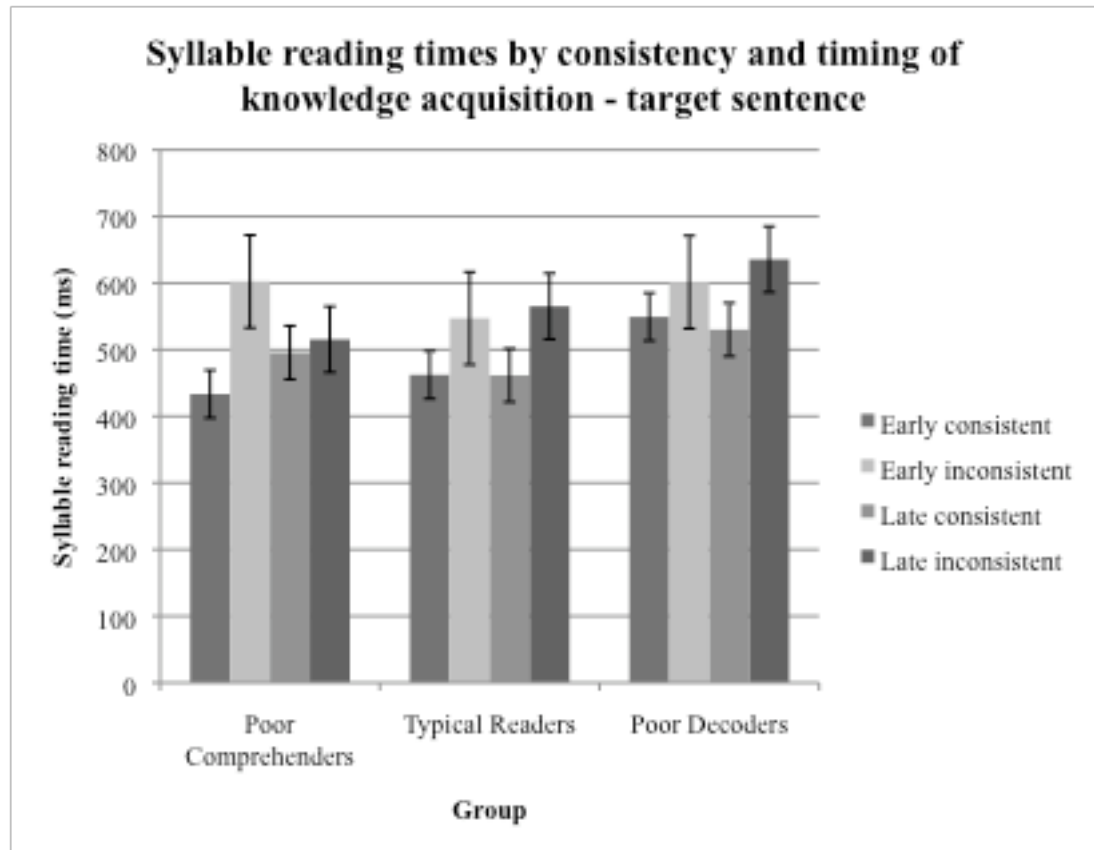


Figure 3.2. Mean syllable reading speeds to target sentences in milliseconds (\* error bars denote 95% confidence intervals)

Mean syllable reading times were entered into a 3 x 2 x 2 Mixed ANOVA. The between-subjects variable was group, which had three levels (poor comprehenders, typically developing children, poor decoders), and the within-subjects variables were consistency, which had two levels (consistent, inconsistent), and familiarity (early-acquired, late-acquired).

The ANOVA revealed a highly significant main effect of Consistency ( $F(1,42) = 21.133, p < .001, \eta^2 = .335$ ). There was no significant main effect of

familiarity ( $F<1$ ) and no significant main effect of group ( $F(2,42)=1.271, p=.29$ ).

There were no significant interactions between consistency and group ( $F<1$ ), familiarity and consistency ( $F<1$ ), and familiarity and group ( $F<1$ ). However, the 3 way Group x Consistency x Familiarity interaction was significant ( $F(2,42)=3.302, p<.05, \eta^2=.136$ ).

As in previous experiments, reading of target sentences was inhibited in all groups when an inconsistency was present. However, the Group x Consistency x Familiarity interaction demonstrated that this effect was less robust in the poor comprehenders. Post-hoc analysis revealed that in fact, poor comprehenders did not show a consistency effect in the late-acquired conditions, whereas both the typically developing and poor decoders showed this effect across all conditions.

### 3.7.2.2 Inhibition Scores

In order to control for individual differences in reading speed, an index of relative inhibition in syllable reading times was calculated for each participant for the early- and late-acquired conditions. Means and standard deviations for the inhibition scores are shown in Table 3.10 below.

*Table 3.10.* Comparison of inhibition in syllable reading times by group

	Poor Comprehenders (N=15)	Typically Developing Children (N=15)	Poor decoders (N=15)
Early acquired	0.5259 (0.675)	0.1981 (0.288)	0.6734 (1.822)
Late acquired	0.0004 (0.221)	0.0263 (0.211)	0.0239 (0.212)

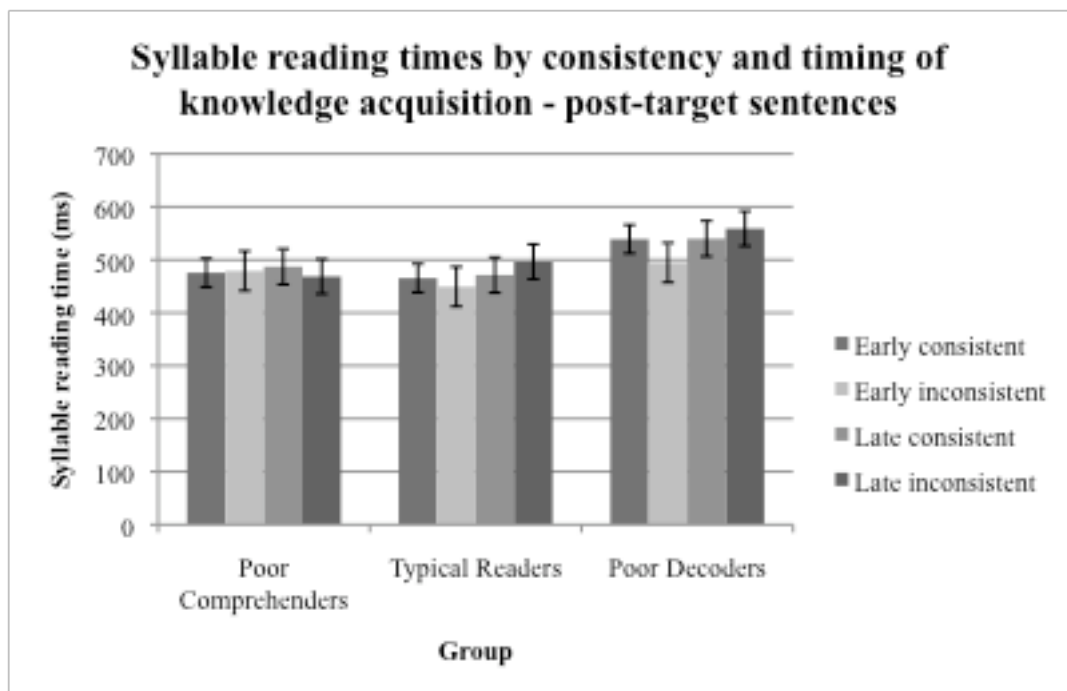
The data were entered into a 3 x 2 mixed analysis of variance. The analysis revealed a significant main effect of familiarity ( $F(1,41)=8.157, p<.01; \eta^2=.17$ ). In line with the reading time results, there was significantly more inhibition of reading in

the presence of inconsistency in the stories that induced inferences about early-acquired information. The main effect of group was not significant ( $F < 1$ ), and nor was the interaction between familiarity and group ( $F < 1$ ). It was noted that the typically developing readers show less inhibition in the early-acquired condition than the poor comprehenders and poor decoders, however this difference was not found to be statistically significant.

### 3.7.3 Reading Times to Post-target Sentences

#### 3.7.3.1 Syllable Reading Times

Mean syllable reading times for each group and condition were computed for the three post-target sentences combined, and these are shown in Figure 3.3 below.



*Figure 3.3.* Mean syllable-reading speeds to three post-target sentences in milliseconds (\* error bars denote 95% confidence intervals)

Mean syllable reading times were entered into a 3 x 2 x 2 Mixed ANOVA. The between-subjects variable was group, which had three levels (poor comprehenders, typically developing children, poor decoders), and the within-subjects variables were

consistency, which had two levels (consistent, inconsistent), and familiarity (early-acquired, late-acquired).

The ANOVA revealed a small but significant main effect of Familiarity ( $F(1,42) = 5.713, p < .05, \eta^2 = .12$ ). In the late-acquired condition, post target sentences were read more slowly. The main effects of group ( $F(2,42)=1.82, p=.18$ ) and consistency ( $F<1$ ) were not significant. Similarly, the interactions between familiarity and group ( $F(2,42)=1.4, p=.26$ ), consistency and group ( $F<1$ ), familiarity and consistency ( $F(1,42)=1.863, p=.18$ ) and the three way interaction between familiarity, consistency and group ( $F(2,42)=1.596, p=.22$ ) were not significant.

### 3.7.3.2 Inhibition Scores

As before, in order to control for individual differences in reading speed, an index of relative inhibition in syllable reading times was calculated for each participant for the early- and late-acquired conditions. Means and standard deviations for the inhibition scores are shown in Table 3.11 below.

*Table 3.11.* Comparison of inhibition in syllable reading times by group

	Poor Comprehenders (N=15)	Typically Developing Children (N=15)	Poor decoders (N=15)
Early acquired	0.0171 (0.1431)	-0.074 (0.1103)	-0.1169 (0.1213)
Late acquired	0.0044 (0.1226)	0.1700 (0.2080)	0.1294 (0.2048)

The data were entered into a 3 x 2 mixed analysis of variance. The analysis revealed a significant main effect of familiarity ( $F(1,41)=8.353, p < .01; \eta^2 = .17$ ). This analysis showed that there was significantly more inhibition of reading in the presence of inconsistency in the stories that induced inferences about the later-acquired information. The Group x Familiarity interaction was marginally significant

( $F(1,41)=2.784, p=.07, \eta^2=.12$ ), and suggested the trend that the poor comprehenders did not show facilitation of reading in the early-acquired condition like the typical readers and poor decoders did, and that in the late-acquired condition they did not show as much inhibition of reading than either the typical readers or the poor decoders. The main effect of group was not significant ( $F(2,41)=1.795, p=.18$ )

### *3.8 Discussion*

Participants read inconsistent target sentences more slowly than consistent target sentences, confirming the hypothesis that an inconsistency presented in the text would lead to inhibition of reading. Unlike in Experiment 3, and therefore contrary to prediction, this effect was not limited to the early-acquired stories in the typically developing children and poor decoders. However, as in Experiment 3, the poor comprehenders did not show an inconsistency effect in the late-acquired passages. Taken together, these data show that inferences requiring integration of later-acquired information are more difficult for poor comprehenders, and if they are too challenging they are not generated on-line contemporaneously during reading.

As predicted, the poor comprehenders were worse at answering the comprehension questions although in this experiment, the difference did not reach statistical significance. All groups found the inferential questions more challenging than the questions requiring factual information, however performance across all questions types indicates that the children were reading for meaning.

The performance of the poor comprehenders on both the reading of target sentences, and answering comprehension questions mirrors the findings of Experiment 3. However, whilst the performance of the typical readers and poor decoders in the early-acquired condition replicated that in Experiment 3, in Experiment 4 they were

able to generate inferences on-line when the information to be integrated was recently acquired as well.

It was clear from both experiments that the early-acquired information was easy for all groups to integrate with the situation model of the story. The finding that poor comprehenders show a reduced consistency effect in the late-acquired condition is in line with the hypothesis that poor comprehenders would find the integration of such knowledge more challenging because either their inferencing skills are less robust, or their ability to access their knowledge-base is impaired, or a combination of these two factors interact.

It is unclear why the typically developing readers and the poor decoders in Experiment 4 found it easier to integrate the late-acquired information than those in Experiment 3. One possibility is that it may reflect differences in the curriculum or in teaching, as Experiment 4 was conducted in a different school year to Experiment 3. Another possibility is that because participants in Experiment 4 were tested in the Spring Term of Year 4, the information required for the late-acquired inferences had been more recently reinforced and was therefore more accessible and easier to retrieve during reading. Participants in Experiment 3 were tested in the Summer Term, therefore a longer time had elapsed between teaching of the material, and the experiment taking place

It is interesting then, that in Experiment 4, only the poor comprehenders found it difficult to generate inferences in the late-acquired condition. There is no evidence to suggest that they had lower levels of background knowledge, and it was confirmed that all participants were taught the relevant background knowledge (on hibernation, metamorphosis and magnetic force) required to generate the late-acquired inferences, in school during the first term of the academic year.

Experiment 4 also examined the reading times to post-target sentences to determine whether there was any carry-over in processing of inconsistent information evident in the reading times to post-target sentences. No consistency effect was observed, in that the syllable reading times to post-target sentences were not significantly different when there was an inconsistency presented in the target sentence. There was, however, a main effect of familiarity. Syllable reading times to the post-target sentences in the late-acquired conditions were significantly longer, suggesting that integration of more recently acquired information requires extended processing.

Analysis of the inhibition scores calculated to account for individual differences in reading speed, revealed that reading of post-target sentences in the late-acquired condition was inhibited by inconsistency, whereas in the early-acquired condition this was not found to be the case.

Surprisingly, examination of the data revealed that reading of post-target sentences was relatively faster in the early-acquired condition in the typical readers and the poor decoders. This may suggest that on encountering an inconsistency with inferred information, they read subsequent sentences with greater urgency in a bid to find an explanation and restore coherence. This may not be evidenced in the late-acquired condition because of the greater task difficulty.

In the late-acquired condition it does appear that in typical readers and poor decoders there is some carry-over of inhibition caused by the inconsistency. However poor comprehenders show only little evidence of a carry-over effect. These data and the data from the target sentence reading are consistent with the view that the inferences required for the inconsistency to be detected in the late-acquired condition may be too difficult for the poor comprehenders to generate on-line.



### *3.9 Summary and Conclusion*

The findings from Experiments 3 and 4 show that typical readers, poor comprehenders and poor decoders automatically integrate prior knowledge with incoming text during reading when that prior knowledge has been known to them for at least 4 years. It is less clear what happens in the case of later-acquired general knowledge. In Experiment 3, participants did not automatically generate knowledge-based inferences when the information needed to make the inference was more recently acquired whereas in Experiment 4, typical readers and poor decoders did. In Experiment 4, in the later-acquired condition there was also some evidence of carry-over of inferential processing in the typical readers and poor decoders. The poor comprehenders did not show any slowing of reading to inconsistent sentences in the late-acquired condition, and they did not show carry-over in the post-target sentences, so it cannot be argued that their failure to show an inconsistency effect in the target sentences was due to a lag in inferential processing.

The poor comprehenders did not automatically integrate recently acquired knowledge with the text to generate inferences on-line. This may reflect difficulty in accessing the relevant knowledge from semantic memory, or may simply result from a lack of awareness of the need to integrate background knowledge with incoming text.

In conclusion, typical readers and poor decoders are able to generate inferences automatically during reading that require prior knowledge to be integrated with incoming text. Poor comprehenders can generate inferences automatically during reading if the information to be integrated has been in the knowledge-base for sufficient time. It seems, however, that they are not able to make use of recently acquired knowledge to generate inferences on-line.

## CHAPTER 4

### **The Online Generation of Emotion-state Inferences and Spatial**

### **Inferences in the Auditory Modality**

#### *4.1 Introduction*

The previous experiments have shown that, during reading, children automatically generate inferences about fictional characters' emotion-states, and about spatial features of objects or locations described in the text. It has also been shown that children can integrate familiar early-acquired knowledge with facts they encounter within a passage but they find it harder to generate inferences online about information that has been acquired later. All of the studies have focused on inference generation during reading, but do not address the hypothesis that the difficulties shown by poor comprehenders are more general and affect such processing during listening too. It seems logical to predict that a connection will exist between listening comprehension and reading comprehension, and published research supports this for typically developing children (e.g. Kendeou, Bohn-Gettler, White & van den Broek, 2008). If one considers oral language comprehension to be an integral part of 'linguistic comprehension' as proposed in the Simple View of Reading (Gough & Tunmer, 1986), then children who show poor reading comprehension in the presence of normal decoding would be expected to have difficulties comprehending oral language. Assuming that linguistic comprehension skills are consistent across different media, and that decoding difficulties constrain reading comprehension, one would predict that poor readers and typically developing children would perform at a similar level in an oral comprehension task.

In the experiment reported in this chapter, the stories used in Experiments 1 and 2 were presented auditorily using self-paced listening paradigm (analogous to that

used for reading), to determine whether emotion-state and spatial inferences are generated on-line during listening as well as during reading. Participants in this experiment were asked to listen to the same short stories, which either a consistent or inconsistent emotion-state ascribed to the protagonist, or contained a consistent or inconsistent spatial detail. In order to notice inconsistencies in the passages it was necessary for participants to have generated particular inferences as they listened to them. It was hypothesized that typically developing children and poor decoders would slow their listening down to a similar degree when encountering inconsistency in a passage, whereas for poor comprehenders there would be significantly less slowing, indicating that they either did not notice, or spent less time attending to the inconsistency.

## *4.2 Method*

### *4.2.1 Design*

A mixed design was used with one between-subjects factor, Group (Typically developing children, poor decoders, poor comprehenders) and two within-subjects factors, Consistency (consistent, inconsistent) and Inference Type (emotion-state, spatial). Each child heard the eighteen stories used in Experiments 1 and 2, split across two testing sessions. Within each story, participants heard a target sentence that related to information that could be inferred from the detail given in the story. In nine sentences, the content of the sentence was consistent with the inferred information, and in the other nine sentences, the content was inconsistent. Five consistent and five inconsistent sentences related to emotion-states of the main protagonists, and four consistent and four inconsistent sentences related to spatial details that could be inferred from prior information given in the stories. Measures were taken of listening

times to the target sentence containing the information requiring children to refer back to a previously generated inference.

#### 4.2.2 *Materials*

##### 4.2.2.1 *Experimental Stories*

The consistent and inconsistent stories used in Experiments 1 and 2 were used unchanged for this experiment, and can be found in full in Appendix 1 (emotion-state stories) and Appendix 2 (spatial inference stories).

Recordings of each story were created using Adobe Audition Version 1.5 software. The stories were read aloud by the female experimenter at a moderate pace, in a clear, neutral accent. Each story was recorded in full initially, and then the files were cut into individual sentences using the editing facility in Adobe Audition. The recordings were created in this way in order to maintain prosodic integrity.

All stories described situations that children of this age group might find themselves in. Towards the end of each story, there was a target sentence that related to information that might reasonably have been inferred from the earlier context. In nine stories the target sentence was consistent with such information and in nine sentences the target sentence was inconsistent with such information. For example, the target sentences below were embedded in context where the consistent emotion state of the protagonist was anger.

Consistent: *Jack was very angry indeed*

Inconsistent: *Jack was very happy indeed*

In the first example, Jack was angry because someone had beaten him in a class spelling test by cheating, and in the second example, Jack's father had said that he was going to paint Jack's bedroom pink against Jack's wishes.

#### *4.2.2.2 Comprehension Questions*

To check that children were listening and attending to the passages and to provide an off-line index of comprehension, four questions were asked after each story – two asked for factual information from the story, and two required participants to generate an inference to obtain the correct answer; one required a coherence inference and one an elaborative inference. Both coherence and elaborative inferences require participants to integrate background knowledge with factual information presented in the text, however in the case of coherence inference, the inference must be generated in order for the passage to be understood. In the case of elaborative inference, although the passage can be understood without the inference, the mental representation of the text and the resulting situation model will be more elaborate. The questions used were the same as those used to assess reading comprehension off-line in Experiments 1 and 2 (see Appendices 1 and 2)

#### *4.2.2.3 Standardised Tests of Reading and Language Ability*

Participants from this experiment also took part in Experiment 4, and the same selection criteria were used. Nonword reading, sight word reading and reading fluency were assessed using the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner & Rashotte, 1999). Reading age equivalents and standard scores were calculated for all participants.

Verbal IQ was measured using Wechsler Abbreviated Scales of Intelligence (WASI) comprising a test of vocabulary and verbal similarities (Wechsler, 1999).

Nonverbal IQ was assessed using Ravens Progressive Matrices, administered as a group task.

Passage Reading Accuracy and Reading Comprehension were assessed using form 2 of the Neale Analysis of Reading Ability (NARA-II) – Second British Edition (Neale, 1997).

#### *4.2.3 Participants*

Three groups of fifteen children were originally recruited to take part in this study. Data from two participants (one poor decoder and one typical reader) was lost due to a data recording error, and the experiment was discontinued in one typical reader due to behavioural issues. Data from thirteen typical readers, fifteen poor comprehenders and fourteen poor decoders was entered into the analysis. The characteristics of these participants are shown in table 4.1.

Table 4.1. Characteristics of participants by group

Measure	Group	Poor Comprehenders (n=15)		Typically developing children (n=13)		Poor decoders (n=14)	
		Mean	SD	Mean	SD	Mean	SD
Chronological Age (months)		104.27	4.68	101.85	3.85	104.21	4.87
NARA(II) Reading Accuracy <sup>1</sup> (Form 2)		100.33	6.91	102.46	5.40	88.00	2.96
NARA (II) Reading Comprehension <sup>1</sup> (Form 2)		84.27	7.72	98.85	4.38	89.55	8.43
Verbal IQ (Wechsler Abbreviated Scale of Intelligence <sup>1</sup> (WASI))		95.87	14.37	109.54	6.56	102.00	11.44
Vocabulary (WASI) <sup>2</sup>		42.07	12.55	51.69	9.29	46.21	11.13
Verbal Similarities <sup>2</sup> (WASI)		51.67	7.54	59.00	6.22	56.43	6.37
Nonverbal IQ <sup>1</sup> (Ravens Progressive Matrices)		101.31	8.28	98.35	10.00	97.54	10.50
Test of Word Reading Efficiency (TOWRE) Sight Word Reading <sup>1</sup>		106.27	12.42	106.54	7.10	94.07	7.41
TOWRE Phonemic Decoding Efficiency <sup>1</sup>		104.67	12.78	106.77	7.67	91.29	4.18
Phoneme Deletion (McDougall Test of Phoneme Deletion, /18)		13.27	3.17	14.54	1.613	11.00	2.57

<sup>1</sup> Standard Score

<sup>2</sup> T-scores

A series of one-way analyses of variance were performed to validate the classification of children into groups. There were no significant between group differences in chronological age and non-verbal IQ. However, there were significant between group differences in Verbal IQ when the two components were combined ( $F(2,39)=4.97, p <.05, \omega^2=.16$ ) and post-hoc testing revealed that the verbal IQ of poor comprehenders was lower than that of the typically developing readers. There was no significant difference between the verbal IQ of the poor decoders and the poor comprehenders. There was a significant main effect of Group in NARA-II reading accuracy ( $F(2,39)=29.06, p<.001, \omega^2=.57$ ), with post-hoc analysis showing that this was accounted for by impaired performance on these tests by the poor decoder group. This was corroborated by the TOWRE; there were significant main effects of Group in both the Sight Word Reading ( $F(2,39)=7.95, p =.001, \omega^2=.25$ ) and the Phonemic Decoding Efficiency ( $F(2,39)=11.84, p<.001, \omega^2=.34$ ) subtests. Post-hoc analysis using Tukey's HSD revealed that the poor decoders were worse than both the typical readers and the poor comprehenders on both subtests again demonstrating their decoding difficulties. There was a significant main effect of group for the results of the McDougall Test ( $F(2,39)=6.61, p <.05, \omega^2=.21$ ), reflecting poorer phonological skills of the poor decoder group.

There was a significant main effect of group on NARA-II reading comprehension ( $F(2,39)=25.99, p <.001, \omega^2=.54$ ), with post-hoc analysis indicating that this was accounted for by superior comprehension in the typical readers. As expected the poor decoders' comprehension was equivalent to that of the poor comprehenders due to it being constrained by their impaired reading accuracy. There was a significant between-group difference in listening comprehension ( $F(2,39)=5.22, p <.05, \omega^2=.17$ ). Post-hoc analysis using Tukey's HSD showed that poor



comprehenders performed significantly worse than the typical readers and marginally significantly worse than the poor decoders ( $p=.06$ ). The results of the listening comprehension test support the hypothesis that the poor decoders' comprehension, as measured by the NARA II Reading Comprehension test, was indeed constrained by their reading ability.

#### *4.2.4. Procedure*

Participants listened to eighteen stories divided equally across two testing sessions. Each of the two sessions lasted approximately twenty minutes. The stories were presented auditorily, sentence by sentence, on a Dell Latitude 120L laptop computer with intrinsic mouse and speakers. The experiment was carried out in a quiet room allocated by the participating schools, away from any classroom noise. Administration of the experiment and collection of reading time data were controlled using E-prime version 1.1 experiment management software. Instructions were presented on the screen and read aloud by the experimenter.

A self-paced listening paradigm was employed. Participants were required to press a mouse button to advance to the next sentence. The computer was set up so that pressing either the left or the right mouse button advanced the story to the next sentence and participants were told they could press either button. The mouse was configured in this way as pilot work had demonstrated that on a few occasions the wrong mouse button was pressed initially, which brought about an artificially inflated reading/listening time as the participant took time to realise they had pressed the wrong button.

Participants were advised to ensure that they only pressed the mouse button once in order to advance to the next sentence. However, on a few occasions, sentences were skipped as the mouse button was pressed more than once. When this

occurred, the experimenter read the skipped sentences to the participant, and a note was made on the test form. If the target sentence was affected then the time recorded was deleted from the analysis.

Listening times were recorded as time between mouse button presses to the nearest millisecond. In order to compare the differences between responses to target sentences in the consistent and inconsistent stories, the time taken for the auditory file to play was deducted from the total for the listening time in order to account for any differences in length of the sound files. The time in milliseconds entered into the analysis therefore reflects time spent on activity other than basic listening to the reading of the passage. At the end of each story, a screen with the word ‘questions’ came up and the experimenter asked four questions to check participants’ comprehension of the story.

### *4.3 Results*

#### *4.3.1 Comprehension Questions*

Participants’ responses to the questions asked at the end of each story are shown in Table 2. Overall, the children performed well answering the questions, indicating that they were listening attentively.

*Table 4.2. Means (and Standard Deviations) of scores to responses to questions asked at the end of each story.*

Question Type	Poor Comprehenders (N=15)	Typically Developing Children (N=13)	Poor decoders (N=14)
Factual (/36)	21.40 (5.08)	26.15 (3.05)	24.00 (3.80)
Inference (/36)	17.93 (4.91)	23.08 (1.80)	20.50 (5.36)

Data from the comprehension question scores were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was question type (factual, inferential). The analysis revealed a significant main effect of question type ( $F(1,39)=46.66, p < .001, \eta^2=.545$ ). Children were better at answering questions that asked for literal information, than those requiring the generation of an inference. There was also a significant main effect of group ( $F(2,39)=5.50, p < .01, \eta^2=.220$ ). Post-hoc analysis using the Tukey-Kramer approach revealed that the poor comprehenders were significantly worse than the typically developing children at answering the comprehension questions. The scores of the poor decoders fell mid-way between those of the poor comprehenders and the typically developing readers. The effect size was medium in all cases; (poor decoders compared with poor comprehenders -  $d=.58$ ;  $d=.50$ ; poor decoders compared with typical readers  $d=.62$ ;  $d=.64$ ). The Group x Question Type Interaction was not significant ( $F<1$ )

Half of the inference questions could be answered correctly if the appropriate coherence inference was generated, and the other half required generation of an elaborative inference. Table 4.3 shows the accuracy of responses to the inference questions by group.

*Table 4.3.* Means (and Standard Deviations) of scores to responses to inferential questions asked at the end of each story.

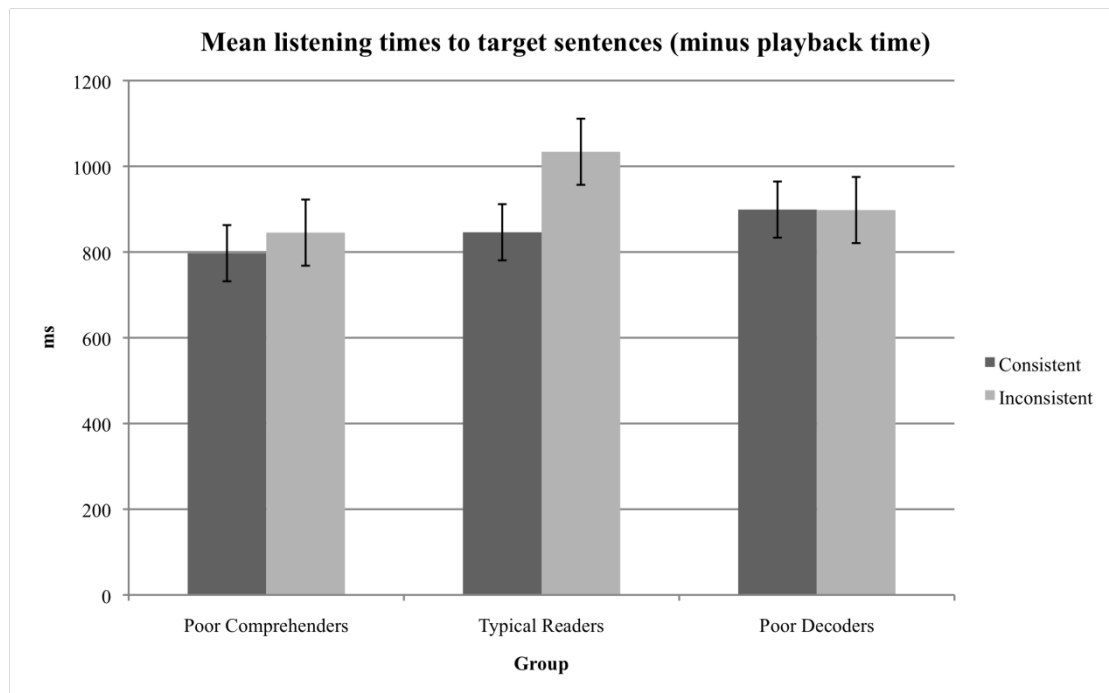
Question Type	Poor Comprehenders (N=15)	Typically Developing Children (N=13)	Poor decoders (N=14)
Coherence (/18)	9.60 (2.29)	12.38 (1.19)	10.29 (2.81)
Elaborative (/18)	8.33 (3.11)	10.69 (1.44)	10.21 (2.89)

Responses to the inference questions were entered into a two way mixed analysis of variance. The between-subjects variable was group (poor comprehenders, typically developing children, poor decoders) and the within-subjects variable was inference type (coherence, elaborative). The analysis revealed a significant main effect of inference type ( $F(1,39)=9.54, p < .01, \eta^2=.197$ ) with the elaborative inference questions proving more difficult for participants than the coherence inference questions. There was also a significant main effect of group ( $F(2,39)=4.80, p < .05, \eta^2=.198$ ), and post-hoc analysis using Tukey's HSD revealed that the poor comprehenders were significantly worse at answering the inferential questions than the typical readers. The poor decoders' performance was worse than that of the typical readers but better than that of the poor comprehenders, however once again, it

was not statistically significantly different from either group. The interaction between question type and group was not significant ( $F(2,39)=2.16, p > .05$ ), however, it was interesting to note that poor decoders were marginally worse than the typical readers at answering the coherence inference questions, and unlike either the poor comprehenders or typical readers, who found elaborative inference questions more difficult, there was little difference between poor decoders' accuracy in answering either type of inference question. The effect sizes suggest that for coherence questions, performance of poor decoders was more like that of poor comprehenders ( $d = .27$  suggests little effect, compared with  $d = .96$  when comparing with typical readers) whereas for elaborative inference questions the opposite pattern was shown with performance more like that of the typical readers ( $d = .21$  indicating little difference) than the poor comprehenders ( $d = .62$  indicating a medium effect).

#### *4.3.2 Listening Times to Target Sentences*

Mean listening times and 95% confidence intervals by condition and group are shown in figure 1. All figures quoted are total listening times less the time taken for the audio file to play, in milliseconds.



*Figure 4.1.* Mean listening times to target sentences minus time taken for the audio to play in milliseconds (*error bars denote 95% confidence intervals*)

Mean listening times were entered into a 3 x 2 Mixed ANOVA. The between-subjects variable was Group (poor comprehenders, typically developing children, poor decoders), and the within-subjects variable was Consistency (consistent, inconsistent). The ANOVA revealed a significant main effect of Consistency ( $F(1,39) = 6.879, p < .05, \eta^2 = .150$ ). The main effect of group was not significant ( $F(2,39) = 1.16, p > .05$ ) but there was a significant Group X Consistency interaction ( $F(2,39) = 3.456, p < .05, \eta^2 = .151$ ). Post-hoc analysis using the Tukey-Kramer approach showed that the presence of an inconsistency caused the typical readers to pause for longer than either the poor comprehenders or poor decoders before moving on to hear the next sentence in the story.

To provide a measure of interference caused by the inconsistency, taking account of individual differences in processing speed, the difference between mean inconsistent and mean consistent listening times was divided by the mean consistent

listening times for each child (mean inhibition). The inhibition score for typical readers (0.245, SD = 0.261) was significantly greater than that for poor decoders (0.021, SD = 0.234) and poor comprehenders (0.071, SD = 0.219).

#### *4.4 Discussion*

Typical readers advanced to hear a subsequent sentence more slowly after hearing target sentences containing an inconsistency than after hearing target sentences consistent with the previous passage content. Poor comprehenders and poor decoders did not show a statistically significant consistency effect, therefore the time between hearing the end of the target sentence and advancing to the next sentence did not vary with consistency of the target sentence. In typical readers and poor comprehenders, these findings are consistent with the results from analogous experiments in the written modality where poor comprehenders showed less inhibition of reading in the presence of inconsistency than typical readers or poor decoders. Thus, poor comprehenders are either more likely to fail to notice an inconsistency or will make less effort to resolve an inconsistency, regardless of modality. In terms of inference generation, the results are consistent with the notion that poor comprehenders generate fewer on-line inferences, and therefore are less likely to detect inconsistencies with information not explicitly given in the passage.

Contrary to expectation, poor decoders did not pause for longer on the target sentence before moving on to the post-target sentence when there was an inconsistency present. This is at odds with their performance on analogous experiments in the written modality, where they showed similar levels of inhibition to inconsistency as typical readers. Furthermore, this finding seems inconsistent with their performance in two off-line verbal comprehension tasks. During the experiment, they performed at a similar level to typical readers and were more accurate than the

poor comprehenders at answering comprehension questions after the story had been heard in its entirety. This was corroborated by their performance on a background measure of listening comprehension administered prior to the experiment, where they also responded as accurately as typical readers, and more accurately than poor comprehenders.

Further analysis of the poor decoders' performance on the off-line measure of comprehension administered after they listened to each experimental story suggested a trend for poor decoders to be worse than typical readers at answering coherence inference questions. One interpretation of this finding is that it is indicative of a difficulty in constructing coherent representations of incoming oral language (and hence may be related to their performance in the on-line task). Moreover, there is evidence that the poor decoders in this study have weaker underlying language skills than those of the typical readers for example, their verbal IQ was lower than that of the typical readers though their nonverbal IQ was in the same range. Participants in this experiment also took part in Experiment 5, which required them to be reading at least at the eight-year level. It has been suggested that children with decoding skills at the lower extremes of the distribution are more likely to have a 'pure' decoding impairment whereas those with more moderate reading difficulties such as in this experiment are more likely to show a mixed profile (Hagtvet, 2003).

There is relatively little published research focusing specifically on oral language comprehension in poor decoders, however the evidence available does lend support to the notion that, whilst their difficulties are primarily in the phonological domain, poor decoders exhibit subtle but persistent broad-based weaknesses in oral language (e.g. Catts, Adlof & Ellis-Weismer, 2006; Snowling, Bishop & Stothard, 2000; Scarborough, 1990).



Hagtvet (2003) investigated the relationship between decoding and oral and written language comprehension skills in 70 Norwegian native speakers. Language skills assessed included reading and listening comprehension (using both a story retelling and cloze task in both modalities), complex syntax, and vocabulary (WISC-R). She found a significant positive relationship between decoding and all of the language skills measured, even when controlling for IQ. Thus, it was concluded that difficulties in learning to read are associated with subtle general language weaknesses affecting both written and oral modalities, however the precise effects of these were found to be contingent on task demands.

Catts, Adlof and Ellis Weismer (2006) conducted a retrospective analysis of 57 poor comprehenders, 98 typically developing readers and 27 poor decoders' reading and language skills, with participants assigned to groups based on their performance in reading assessment conducted in 8<sup>th</sup> Grade. They analysed assessment data collected when these children were in Kindergarten, 2<sup>nd</sup> Grade and 4<sup>th</sup> Grade. Poor decoders' listening comprehension assessed using an off-line task was equivalent to that of the typically developing children in 2<sup>nd</sup> and 4<sup>th</sup> grades, but it was significantly poorer when assessed in Kindergarten. In a similar vein, Wise, Sevcik, Morris, Lovett and Wolf (2007) found that listening comprehension predicted significant unique variance in word identification skill.

If it is the case that poor decoders have subtle underlying language weaknesses, the speed of the spoken language input in the listening task may have exceeded their processing resources, hence making it difficult for them to engage in the same amount of inferential processing as they would if they were reading. In an on-line reading task, the participant has control over the speed of the language input by virtue of his/her reading speed, and poor decoders read more slowly than typical

readers or poor comprehenders. Reading stories at their own pace may therefore give the poor decoders the additional time that they need to construct elaborate representations of the situation depicted.

An alternative explanation for poor decoders' weak performance on an on-line oral comprehension task is that they are less attentive when listening to stories (a passive task) than when reading them. Indeed, It is well documented that there is considerable co-morbidity between reading disorders and attention disorders (e.g. Gilger, Pennington and DeFries, 1992). Speculatively then it could be that their attention was better engaged when they were required to actively answer questions in the off-line task. Finally, it is possible that when reading text, poor decoders who detect an inconsistency may engage in more attempts to resolve it as soon as it is detected, before moving on. In the case of the reading time experiments reported in this thesis, the poor decoders were not able to engage in re-reading of the prior text but could re-read the target sentence before pressing the mouse button to move on. It is suggested that poor decoders are likely to engage in more re-reading of the text due to a lack of confidence in their own ability to read accurately. They may also make greater efforts to compensate for their reading difficulties by making greater use of their intact comprehension skills to detect and/or resolve inconsistencies with context, such as monitoring comprehension thoroughly and making inferences wherever possible.

In summary, Experiment 5 has shown that typical readers pause when they hear information that is inconsistent with oral passage content. In contrast, poor comprehenders and poor decoders do not pause for additional time when they encounter an inconsistency in oral language that contradicts an inference they have generated on-line during listening. The results in the typical readers and the poor

comprehenders are consistent with the results of Experiments 1-4, suggesting that poor comprehenders' difficulties with inference generation and comprehension monitoring are not limited to the written modality. This experiment therefore provides a replication of the findings of Experiments 1 and 2 in the oral domain in typical readers and poor comprehenders. However, for poor decoders, performance in the oral modality is discrepant with that in the written modality. When poor decoders read a text containing an inconsistency, they slow down in an attempt to resolve it, whereas if they hear a similar text containing an inconsistency they do not. However, in off-line comprehension tasks that are not time-limited, their accuracy appears similar to the typical readers and superior to that of poor comprehenders.

## CHAPTER 5

### **Reading Disability, Passage Coherence and Sensitivity to Context**

#### *5.1 Introduction*

Experiments 1-5 have shown that, during reading and listening, children can automatically generate inferences about fictional characters' emotion-states, and about spatial features of objects or locations described. These experiments have also identified performance differences in poor comprehenders that support the assertion that poor comprehenders have difficulty with inference generation. The results of Experiments 3 and 4 suggest that on-line inference generation is less robust when passages contain information that is less integrated with the reader's general knowledge base. Furthermore, the ability of poor comprehenders to make inferences is fragile and appears to break down more easily when material becomes more challenging.

##### *5.1.1 Standard for Coherence and Inference Generation*

A compelling explanation of poor comprehenders' difficulties with text comprehension is that they adopt and apply a lower standard for coherence than typically developing readers (e.g. Perfetti, Landi & Oakhill, 2007); in turn, this could have a direct influence on their propensity to generate inferences during reading.

Standard for coherence refers to a reader's criterion for accepting comprehension as being adequate or complete. In order for reading to be effective, the reader needs be motivated to acquire a full understanding of the meaning of the text they are reading. It is generally believed that a high standard for coherence fuels higher- level comprehension processes such as comprehension monitoring and inferencing directly. In the case of inference generation, a larger number of inferences will need to be generated in order to meet a high rather than low standard for

coherence unless the text itself is very coherent. Contrary to the traditional typological views of inference generation that make predictions about which inferences are made on-line, inferences are viewed as tools used for coherence maintenance or achievement (Perfetti, Landi & Oakhill 2007). It is argued that inferential activity continues until the reader achieves their standard for coherence and accepts that text comprehension is complete. Crucially, this model makes no predictions about the specific types of inferences that will be generated on-line, rather it takes the view that the standard for coherence operates like a setpoint or criterion for understanding, that indirectly determines the number and variety of inferences generated automatically during reading, as inference generation is halted once a criterion level of coherence is achieved. It is plausible that a number of environmental factors may lead to the adoption of a low standard for coherence, for example low level oral language weaknesses may lead an individual to become accustomed to a poor level of understanding or lack of coherence in the language they experience, and they may therefore fail to realize that richer understanding is available to them.

In support of the notion that inference generation is motivated by a search for coherence, van den Broek, Risen and Husebye-Hartmann (1995) reported three studies investigating how text constraints and standards for coherence interact to drive inference generation in adult readers. In these experiments, an antecedent was defined as necessary when it was essential for the consequence to occur, whereas it was defined as sufficient if the antecedent was likely to be followed by the consequence in the given context, but was not essential for it to occur. They examined the influence of both necessity and sufficiency of antecedents in explaining consequences, on generation of forward and backward causal inferences. In the first analysis reported, they found readers made more extensive use of backward inferences in order to

establish causality when the antecedent was sufficient, but forward inferencing was influenced more by necessity, in that it was more likely to occur if the antecedent was necessary for the consequence. Thus, less backward inferential activity was required in the case of a necessary antecedent for coherence, and more forward inferencing occurred when its antecedent predicted a consequence more reliably.

In their second study they used a story continuation task and manipulated the sufficiency provided for a target inference to see whether it influenced readers' propensity to generate that inference in their story continuation. They found that in the high sufficiency story stem there was a higher probability of readers making the target inference, and readers made few non-target inferences. This provides support for the idea that forward inferencing does occur and is used as a way of checking coherence.

In the third experiment, readers were asked to read sentence pairs and to write a statement that would fit between the two sentences. The extent to which the first sentence was sufficient explanation for the consequence described in the second sentence was varied. The majority of the backward inferences produced in the sentences written were causally related to the events in the second sentence. Sentences written by participants were not causally related to the consequence in two situations; firstly when the first sentence provided a totally sufficient explanation for the second sentence, and secondly when the first sentence did not provide any explanation for the second sentence. Where the second sentence was completely explained there was no need for further causal inference generation, and where there was no causal explanation provided in the antecedent, there was no causal explanation in the sentence generated between that and the consequence. The results here

illustrate how a reader's standard for causal coherence and textual constraints interact to determine inference generation.

#### *5.1.2 Comprehension Monitoring and Standard for Coherence*

Comprehension monitoring is influenced by the standard for coherence adopted. A study by Hacker (1997) found that low skill readers engaged in less monitoring activity than skilled readers, and this was not enhanced by explicit instruction to monitor incoming text for meaning. It was argued that these readers had adopted a lower standard for coherence, and therefore did not see an area for improvement when instructed to monitor meaning. Therefore a low standard for coherence can lead to inconsistencies and ambiguities being ignored or missed (Hacker, 1997). Comprehension monitoring will only lead to detection and resolution of inconsistencies or ambiguities that violate the adopted standard for coherence.

#### *5.1.3. Understanding story structure and standard for coherence*

Cain and Oakhill (1996) asked skilled and less skilled comprehenders to tell a story based on a title prompt. On analyzing the story content they found the less skilled comprehenders produced stories that were less complex in structure and less globally coherent. Cain (1996) also found differences in poor comprehenders' ability to describe features of stories, and what they typically contain. Their descriptions suggest that they were less aware of the functions of story titles, beginnings and endings than their skilled peers. Taken together, the findings of these two studies suggest that a low standard for coherence may permeate into production tasks and may affect understanding of what constitutes a well-structured story.

#### *5.1.4. Rationale for Experiment 6*

In Experiment 6, one plausible account of why poor comprehenders might find it more difficult to generate inferences on-line was explored. Experiment 6 was

devised to determine whether poor comprehenders adopt a lower standard of coherence during reading, and therefore accept greater discrepancies within the text than typically developing or poor decoders. If poor comprehenders are presented with answer options to evaluate, if they adopt a lower standard for coherence then they are likely to make their responses more quickly because they will spend less time evaluating the implications of each answer.

In this experiment, participants were presented with a set of items, consisting of a scenario, a question, and four possible responses varying in plausibility that they were required to accept or reject. In order to generate an answer to the question, participants would need to make inferences about the scenario described. In order to evaluate each answer presented, they would need to decide whether the response could be consistent with an inference that could reasonably be generated from the scenario. The responses participants evaluated varied in plausibility, both in terms of how coherent they were with the scenario (based on real world knowledge), and how they fit with the specific context of the passage. One of the four answers was correct (plausible and coherent), one was plausible in the given context, but less coherent, one was less plausible in the given context, and less coherent than the correct answer, and one answer was incorrect (implausible and incoherent). The valence of participants' responses and the response times were analysed to determine whether there were any observable between-group differences. Firstly, it was hypothesized that poor comprehenders would show faster response times than either poor decoders or typical readers, indicating that they spend less time evaluating the implications of each response. Secondly, when faced with uncertainty about whether an item was plausible or not, it was hypothesized that poor comprehenders' would be more likely to affirm that a response might be correct, due to acceptance of a lower standard for coherence.



## 5.2 *Experiment 6 Method*

Thirty-two test scenarios and 4 practice items were presented to participants divided equally over two testing sessions. Participants were asked to read aloud each scenario and question. On completion of the reading, the examiner pressed a key, the scenario and question disappeared from the screen and four possible answers appeared one by one. In order to advance to the next answer, participants had to indicate whether they thought each answer was plausible by pressing either Y or N on the keyboard.

### 5.2.1 *Design*

A mixed design was used with one between-subjects factor, Group (Typically developing children, poor decoders, poor comprehenders) and one within-subjects factor, Answer type (Correct, Plausible coherent, Plausible incoherent, Incorrect). Each child read 32 short scenarios and 4 practice scenarios divided across two testing sessions. Each scenario consisted of 2-4 sentences that described a familiar scenario. In each case this was followed by a question, with 4 possible answers. The answers to the questions could only be inferred from the scenarios and were never explicitly stated. The four possible answers were presented one at a time in pseudorandom order, and participants were required to respond either yes or no to indicate whether they thought each answer could be right. They were asked to respond 'yes' to any answers that they felt either 'might be right' or were 'definitely right', and only respond 'no' to an answer that was definitely wrong. The scenarios were concerned with aspects of everyday life that would be familiar to children of this age, such as riding a bicycle or eating a meal out.

### *5.2.2 Materials*

#### *5.2.2.1 Item Generation*

Forty-five potential test items selected from a pool of scenarios and questions were put forward for further development. The scenarios created were designed to reflect situations that children in Years 3 and 4 would easily understand. In order to generate responses that varied in plausibility, fifteen native English-speaking undergraduates were recruited, and asked to provide four answers to each question that varied in likelihood of being correct. They were asked to provide answers that were very likely, quite likely, quite unlikely and very unlikely to be correct. The most frequently occurring answer for each was used for the test items. Details of the items and the instructions given to participants are provided in Appendix 5.

The most frequently occurring answers generated for each question and response level were evaluated further. In order to validate the plausibility of the items selected, a rating study was carried out. Thirteen native English speaking undergraduate students were asked to rank the answers provided to the questions in order of plausibility. Details of the items and instructions given to participants are provided in Appendix 5. The thirty-two items with responses that were most consistently ranked were used as the test items. The four next most consistently ranked were used as practice items.

#### *5.2.2.2 Experimental Scenarios*

Two sets of 18 scenarios (16 test items plus 2 practice items) were generated via the procedure outlined above. Each scenario was 2-4 sentences in length, and very easy to read for children of this age group. Scenarios contained between 16 and 33 words, and the mean number of words per scenario was 24 ( $sd = 4.96$ ). The scenarios depicted were events or occurrences that children would be familiar with. Each

scenario was followed by a question; the answer to each question could reasonably be inferred from the scenario presented. Crucially, in order to answer the question inferences would need to be made, as the information required to answer the question was not explicitly given in the passage. Four alternative answers were presented, which varied in terms of both plausibility and coherence with the scenario. Each set of four answers contained one correct answer, (both plausible and coherent), one plausible but less coherent answer, one less plausible and less coherent answer and one incorrect answer (implausible and incoherent). Two example scenarios together with questions and responses are shown in Table 5.1.

*Table 5.1.* Example scenarios with question and possible responses

<b>Example Item 1:</b>	
<b>Scenario:</b> John rode as fast as he could, but could not make it up the steep hill. He stopped for a rest, then he pushed it the rest of the way to the top.	
<b>Question:</b> What was John riding?	
<b>Answer type</b>	<b>Answer presented</b>
Correct (plausible & coherent)	Bicycle
Plausible, less coherent	Motorbike
Less plausible, less coherent	Horse
Incorrect (implausible & incoherent)	Tractor
<b>Example Item 2:</b>	
<b>Scenario:</b> The car was making a terrible noise. Ryan decided it was not safe to carry on driving. He pulled over and rang for help.	
<b>Question:</b> Who did Ryan ring?	
<b>Answer type</b>	<b>Answer presented</b>
Correct (plausible & coherent)	Breakdown company
Plausible, less coherent	Garage
Less plausible, less coherent	Policeman
Incorrect (implausible & incoherent)	Grandma

Reading ability required to read each scenario was assessed using the Hatcher book grading system (Hatcher, 2000). According to this system, children could read all scenarios easily with a reading age of 7 years and above. The scenarios and details of the grading and reading ages required for each scenario are provided in full in Appendix 5.

### *5.2.3 Participants*

Three groups of fourteen children took part in this study. Characteristics of the children taking part in this experiment are shown in table 5.2 below. These children also took part in Experiment 4 and were selected and allocated to groups according to the same criteria outlined in Chapter 3 of this thesis.

Table 5.2. Characteristics of participants by group

Group Measure	Poor Comprehenders (n=14)		Typically developing children (n=14)		Poor decoders (n=14)	
	Mean	SD	Mean	SD	Mean	SD
Chronological Age (months)	104.07	4.80	102.43	3.92	104.5	4.78
NARA (II) Reading Accuracy (Form 2) <sup>1</sup>	100.64	7.07	102.36	5.54	87.71	3.15
NARA (II) Reading Comprehension (Form 2) <sup>1</sup>	83.93	7.90	99.29	5.01	86.50	3.84
Verbal IQ (Wechsler Abbreviated Scale of Intelligence (WASI) <sup>1</sup>	95.71	14.90	110.21	7.08	100.43	12.05
Vocabulary (WASI) <sup>2</sup>	41.71	12.95	51.93	9.02	45.29	9.90
Verbal Similarities (WASI) <sup>2</sup>	51.79	7.81	59.64	6.44	55.21	7.02
Nonverbal IQ (Ravens SPM) <sup>1</sup>	101.77	8.43	99.47	10.11	96.19	12.10
Listening Comprehension <sup>1</sup>	88.73	11.88	102.53	14.76	97.80	10.31
TOWRE Sight Word Reading <sup>1</sup>	106.14	12.88	106.79	7.47	95.14	6.20
TOWRE Phonemic Decoding Efficiency <sup>1</sup>	104.71	13.26	106.79	8.31	91.57	4.22
Phoneme Deletion (McDougall Test of Phoneme Deletion)	13.43	3.23	14.07	2.02	10.86	2.57

<sup>1</sup>Standard Scores <sup>2</sup>t-scores

A series of one-way analyses of variance were performed to check the matching of groups on chronological age, NARA-II reading accuracy, reading

comprehension and IQ. There were no significant between group differences in chronological age, or nonverbal IQ. However, there were significant between group differences in Verbal IQ ( $F(2,39)=5.51, p <.01, \omega^2=.18$ ) and post-hoc testing revealed that the poor comprehenders' verbal IQ was lower than that of the typically developing readers. The difference in verbal IQ was evident in both the vocabulary ( $F(2,39)=3.25, p <.05, \omega^2=.10$ ) and the similarities ( $F(2,39)=4.30, p <.05, \omega^2=.14$ ) subtests that make up Verbal IQ. There was no significant difference between the IQ of the poor decoders and the poor comprehenders, or between that of the poor decoders and the typical readers. There was a significant main effect of group in NARA-II reading accuracy ( $F(2,39)=29.72, p <.001, \omega^2=.58$ ), with post-hoc analysis showing that this was accounted for by impaired performance on these tests by the poor decoder group. There was a significant main effect of Group on NARA-II reading comprehension performance ( $F(2,39)=27.79, p <.001, \omega^2=.56$ ), with post-hoc analysis indicating that this was accounted for by superior comprehension in the typical readers. It is assumed that the poor decoders' comprehension was equivalent to that of the poor comprehenders due to their impaired reading accuracy. This assumption is validated by the results from the Listening Comprehension test; there was a significant main effect of Group ( $F(2,39)=4.44, p <.05, \omega^2=.16$ ), and post-hoc analysis revealed that the poor comprehenders were significantly worse than the typical readers on this task. The poor decoders' performance was similar to that of the typical readers on this task, although the difference between their performance and that of the poor comprehenders did not reach statistical significance. The results indicate that selection and matching of groups was adequate. Further evidence of this is provided from the results of the TOWRE and McDougall Test of Phoneme Deletion. There were significant between-group differences in both the Sight Word

Reading ( $F(2,39)=6.92, p <.01, \omega^2=.22$ ) and the Phonemic Decoding Efficiency ( $F(2,39)=10.86, p <.001, \omega^2=.32$ ). Post-hoc analysis using Tukey's HSD revealed that the poor decoders were worse than both the typical readers and the poor comprehenders on both subtests. A similar pattern of results was observed in the McDougall Test of Phoneme Deletion, in which poor decoders obtained significantly lower scores than either poor comprehenders or typical readers ( $F(2,39)=5.76, p <.01, \omega^2=.18$ ).

#### *5.2.4 Procedure*

The experiment was conducted using a Dell Latitude 120L laptop computer with intrinsic mouse. Administration of the experiment and collection of response data were controlled using E-prime version 1.1 experiment management software. Instructions were presented on the screen and read aloud by the experimenter.

Participants were presented with four practice and thirty-two experimental scenarios divided equally between two testing sessions. Each of the two sessions lasted approximately fifteen minutes.

For each test item, the scenario and the corresponding question appeared on the computer screen. Participants were instructed to read the scenario and the question aloud. Immediately after the question had been read, the experimenter pressed a key that triggered the scenario and question to disappear from the screen and the possible answers to appear. Four possible answers were presented to the participants one at a time. To control for order effects, the four answers were presented in a pseudorandom order, ensuring that participants received each condition in each position the same number of times (i.e. They got the 'correct' answer presented first, second, third and fourth an equal number of times). Participants were asked to respond to each answer by pressing Y on the keyboard if they thought the answer was correct or might be

correct, or N if they thought the answer was definitely incorrect. It was emphasised to participants that it was acceptable and expected that they would press Y for more than one of the four possible answers for each question. For the purpose of the experiment “B” on the standard computer keyboard was configured to indicate a Yes response, and was labelled as such, so that Y and N were next to each other on the keyboard. Participants’ responses triggered the next answer to appear on the screen. Once participants had responded to each of the four possible answers, the next scenario was presented on screen. The sixteen experimental items administered in each testing session were presented in randomised order. The order of sessions 1 and 2 was also randomised so that participants did not always get the same 16 items first and last.

### *5.3 Results*

#### *5.3.1 Yes responses*

Numbers of yes responses for each condition are shown by group in Figure 5.1.



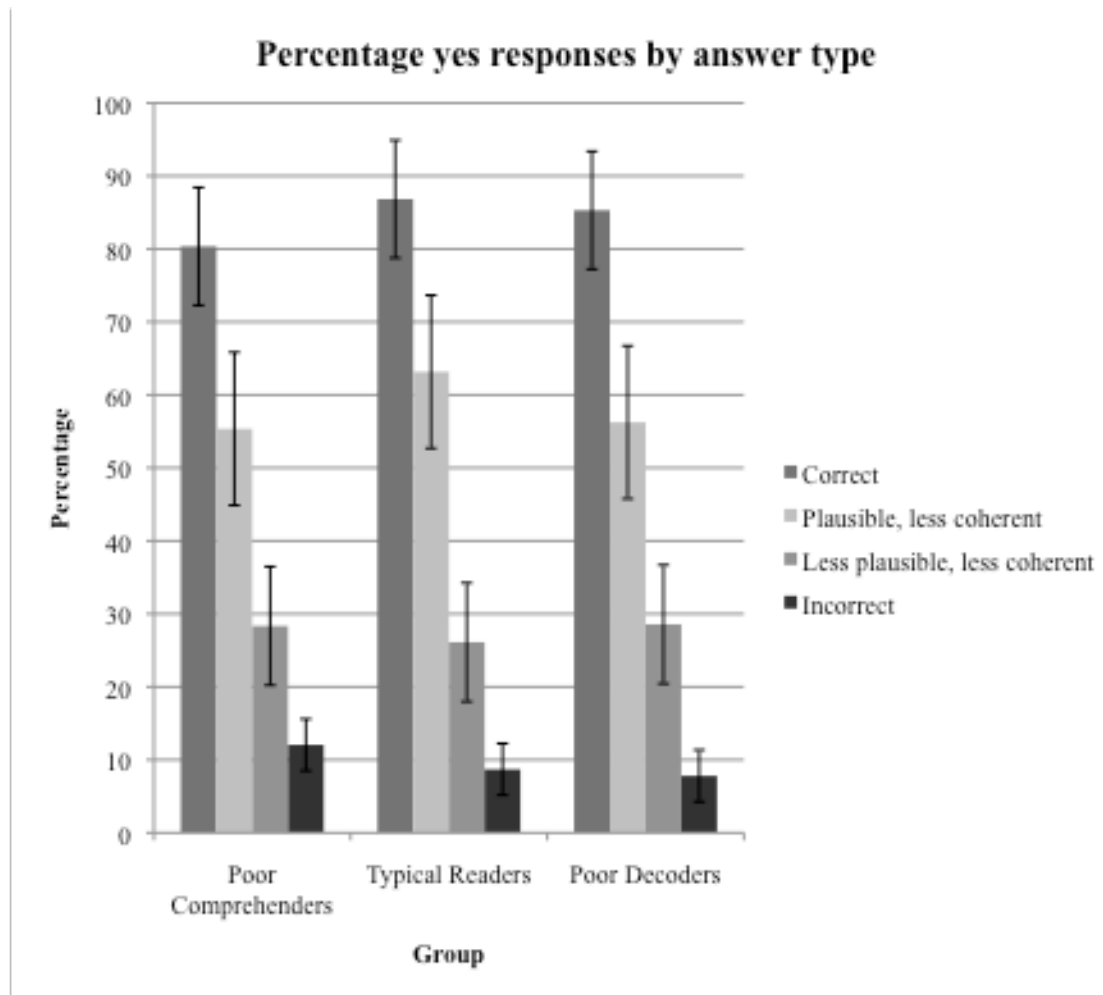


Figure 5.1 Percentage 'yes' responses for each answer type by group

(\*error bars denote 95% confidence intervals)

Numbers of yes responses for each answer type were entered into a 3 x 4 mixed analysis of variance. The between-subjects variable was group (Typical Readers, Poor Comprehenders, Poor Decoders) and the within-subjects variable was answer type (Plausible; Plausible, less coherent; Less plausible, less coherent, and Implausible). The analysis revealed a large and significant main effect of answer type ( $F(3,117)= 423.61, p <.001, \eta^2=.92$ ). The number of 'yes' responses went down in a stepwise fashion with most 'yes' responses in the plausible condition, then plausible & less coherent, then less plausible & less coherent, and the lowest number of 'yes' responses was obtained to the implausible condition. Neither the main effect of group

( $F < 1$ ) nor the group by answer type interaction ( $F(6,117) = 1.20$ ,  $p = .31$ ) reached significance.

### 5.3.2 Response Times

Mean response times for each of the four answer types were calculated for each group. Response times were included regardless of valence of response. Response times are plotted for each answer condition, and are shown by group in Figure 5.2.

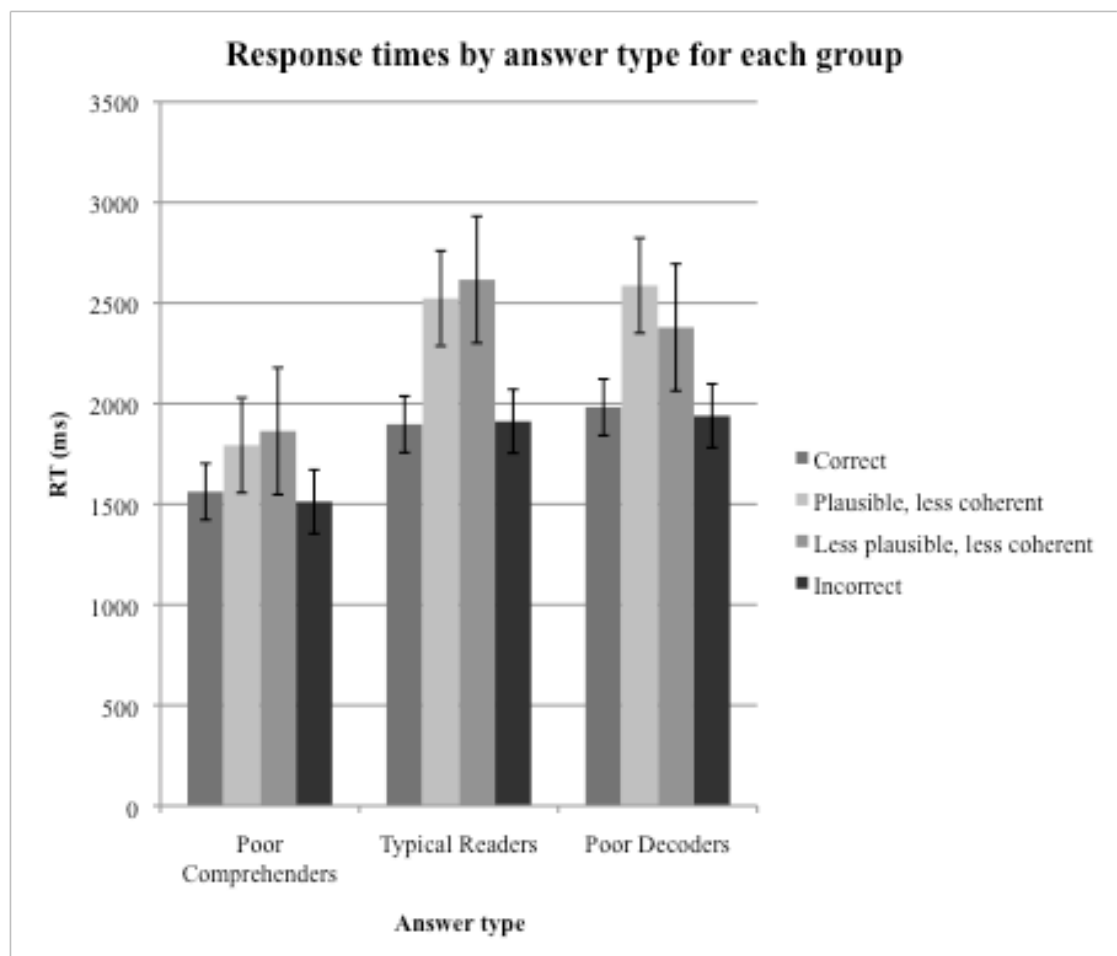


Figure 5.2 Mean response times in milliseconds to each answer type as a function of group. (\*error bars denote 95% confidence intervals)

Mean response times for each answer type were entered into a 3 x 4 Mixed Analysis of Variance. The between-subjects variable was Group (Typical Readers, Poor Comprehenders, Poor decoders) and the within-subjects variable was Answer

type (Plausible; Plausible, less coherent; Less plausible, less coherent; Implausible). The analysis revealed a significant main effect of Answer type ( $F(3,117)=17.65$ ,  $p < .001$ ,  $\eta^2 = .31$ ). Post-hoc analysis revealed that responses to the plausible, less coherent and the less plausible, less coherent answers were made significantly more slowly than to the plausible and implausible answers. The analysis also revealed a significant main effect of Group ( $F(2,39)=3.673$ ,  $p < .05$ ,  $\eta^2 = .16$ ). Post-hoc analysis revealed that poor comprehenders responded faster to all of the answers presented than either the typical readers or the poor decoders. The group x answer type interaction did not reach significance ( $F(6,117)=1.22$ ,  $p = .30$ ).

### *5.3.3 Response Times to Uncertain Answers*

It was noted that the responses to answer conditions 2 (Plausible, less coherent) and 3 (Less plausible, less coherent) were made more slowly than to the more definite answer types 1 (Plausible) and 4 (Implausible), and response times were more variable. Data from these responses was analyzed further to determine whether there were any differences in the time taken for a participant to generate a 'no' response over the time to generate a 'yes' response, and whether there were any between-group differences in these response times. The overall response times for 'yes' and 'no' responses to answer types 2 and 3 are presented in figure 5.3 and are broken down by group in figure 5.4.

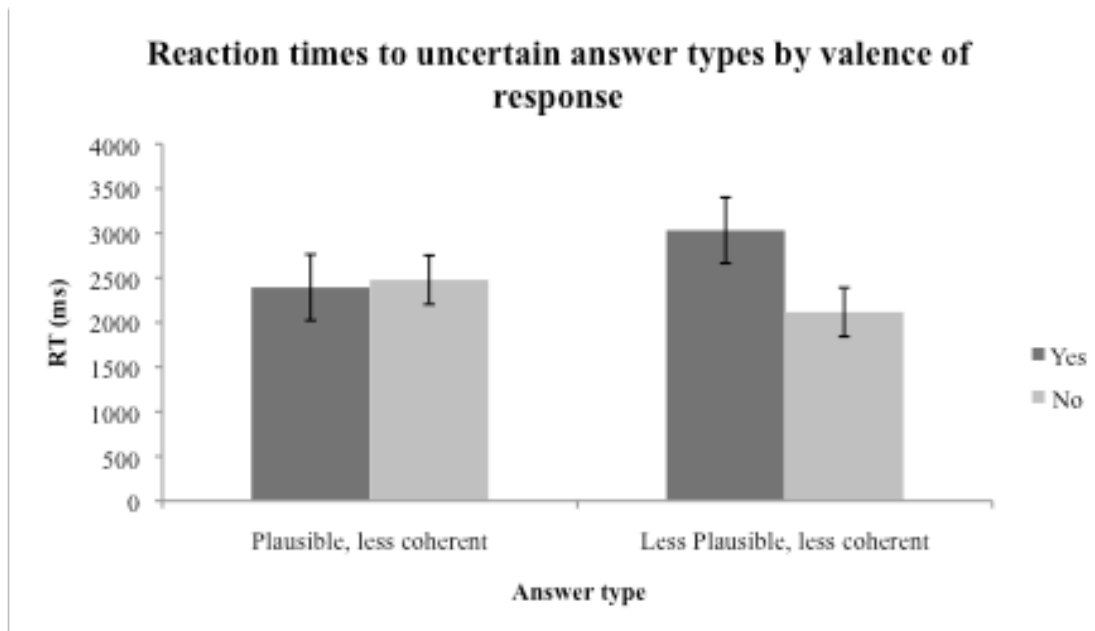


Figure 5.3 Reaction times of responses to uncertain answers by valence of response (\*error bars denote 95% confidence intervals)

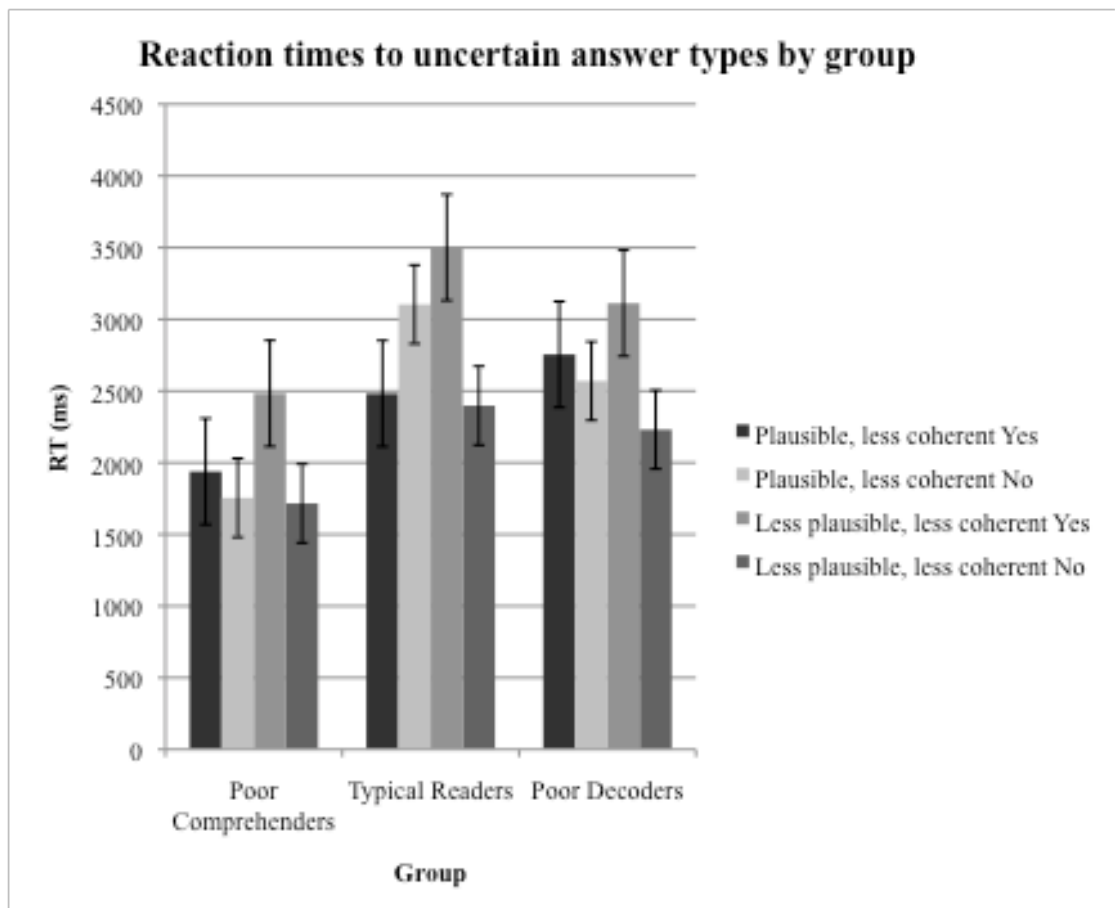


Figure 5.4 Reaction times to responses in uncertain conditions by group (\*error bars denote 95% confidence intervals)

Data were entered into a 3 way mixed analysis of variance. The between-subjects variable was group (Typical readers, Poor comprehenders, Poor decoders) and the within-subjects variables were Answer type (Plausible, less coherent; Less Plausible, less coherent) and Response Valence (Yes, No). The analysis revealed a significant main effect of group ( $F(1,39)=3.413, p < .05$ ). There was also a significant main effect of Valence ( $F(1,39)=8.921, p < .01, \eta^2=.19$ ), and a significant answer type x valence interaction ( $F(1,39)=16.171, p < .001, \eta^2=.29$ ). Post hoc analysis using Tukey's HSD revealed that the poor comprehenders' responses were faster to make their response choices than either the typical readers or the poor decoders. Participants took much longer to make 'yes' responses overall and post hoc analysis showed that both the main effect of answer type and the interaction between plausibility and valence were accounted for by participants taking much longer to make 'yes' responses in the 'less plausible, less coherent' condition. The main effect of answer type was not significant ( $F(1,39)=1.563, p=.22$ ). The interactions between answer type and group ( $F<1$ ), valence and group ( $F<1$ ) and the three-way interaction between answer type and valence and group ( $F(2,39)=2.13, p=.13$ ) were not significant.

#### *5.4 Discussion*

In Experiment 6, typical readers, poor decoders and poor comprehenders were equally accurate at evaluating answers given to questions about short scenarios. As predicted, poor comprehenders were faster at responding, suggesting that they spent less time evaluating all possible ramifications before deciding upon a response.

These results are consistent with the view that poor comprehenders implement a less rigorous standard for coherence when dealing with text. However, a particularly surprising finding was that despite spending less time considering their response, the

poor comprehenders were just as accurate as typical readers and poor decoders across all conditions. There are a number of plausible explanations that may account for this result. First, the task was very easy, requiring only very simple decoding and comprehension of short and straightforward situations that would be well established within participants' general knowledge base. Second, given that they were so short, it is plausible that the scenarios presented were only testing standards for local coherence and were not long enough to test standards for global coherence, which are more likely to vary. In Experiments 1-4, participants were required to integrate premises over longer distances in text, which provided a greater test of poor comprehenders' inferential skills, and results showed that they made fewer inferences, which would be consistent with a lower standard for global coherence. Furthermore, poor comprehenders have been shown to have proportionally greater difficulty on a number of tasks when information needs to be integrated across longer passages of text such as anaphor resolution (Yuill & Oakhill, 1991, 1988), inferring novel word meanings from context (Cain, Oakhill & Lemmon, 2004; Cain, Oakhill & Elbro, 2003), and detecting inconsistencies (Oakhill, Hartt & Samols, 2005). Catts, Adlof and Ellis Weismer (2006) found that poor comprehenders were worse than typical readers at answering inference questions about aurally presented passages if they were required to integrate premises located more than four sentences apart in the text, but performed normally if the information was located in adjacent sentences. Thus, there is evidence that poor comprehenders can generate inferences like typical readers providing pieces of information requiring integration are in close proximity in the text. Importantly, they perform the same as typical readers when answering literal questions, therefore it cannot be argued that this reflects poor memory for the text. It is also worth mentioning that none of the published evidence looking at distance

between premises requiring integration contain any timed component, therefore it is possible that poor comprehenders may also take less time before giving an answer when questioned.

A third possibility is that poor comprehenders were able to perform this task well because demands of the comprehension task were explicitly set out for them to comply with. In contrast, in normal reading automatic inferencing is an implicit task. In a study of comprehension in readers with advanced decoding skills relative to their verbal ability, Snowling and Frith (1986) found that children with low verbal ability with and without autism spectrum conditions performed poorly on an error detection task that required them to identify words inconsistent with story or sentence context, yet they performed much better on an analogous gap-filling task, where they were asked to select from implausible, sentence-appropriate and story-appropriate alternatives. Thus, if they were explicitly given the materials to consider throughout the task, with instructions, they were able to perform the task, but could not perform an analogous task automatically during reading. In essence the nature of the gap-filling task meant that monitoring levels were kept consistently high, whereas comprehension monitoring was poor when under participants' own control. It is worth noting also that poor comprehenders tested here had lower verbal ability relative to controls, while non-verbal ability was the same in both groups. It is suggested that the poor comprehenders performance in Experiment 6 may have been enhanced by the fact that they were explicitly required to evaluate and respond to each possible answer individually.

It is clear that further experiments are needed in order to ascertain whether poor comprehenders' superior performance in this task can be attributed to the task

being very easy, the length of the scenarios presented, or the explicit nature of the task itself. Given that poor decoders' performance was equivalent to that of the typical readers, it would be useful to run the experiment again with just the typical readers and poor comprehenders which would enable longer vignettes requiring more advanced decoding skills to be used. If the hypothesis that poor comprehenders' faster responding is due to them applying a lower standard for coherence holds, one would expect that their accuracy would suffer if the task demands were increased.



## CHAPTER SIX

### General Discussion

#### *6.1 The Simple View and Experimental Findings*

The aim of this thesis was two-fold, first to investigate children's ability to generate different kinds of inferences, and second, to explore factors that might account for individual differences in inferencing skill. Using the Simple View of Reading (Gough & Tunmer, 1986) as a framework, the main experimental paradigm was a self-paced reading task, with a listening task introduced in Experiment 5 to explore possible modality-specific aspects of performance. The final experiment explored the concept of 'standard for coherence' as a potential explanation for individual differences in inferencing skills. Importantly, all of the experiments compared the performance of poor decoders (indexed by poor word-level reading skills), poor comprehenders (indexed by poor comprehension in the presence of age-appropriate decoding) and typically developing readers (age-appropriate decoding and comprehension). This design has the potential to elucidate both deficient and intact component reading skills and yet has to date rarely been used to investigate individual differences in reading skills.

The Simple View of Reading proposes that both decoding and linguistic comprehension skills are required for successful reading comprehension. Both of these skills operate independently, and neither is sufficient to guarantee fluent reading for meaning. This model leads to predictions about how reading comprehension will be affected if one of the underlying component skills is impaired: In pure reading and comprehension disorders, dissociation between decoding ability and linguistic comprehension would be predicted. Further, according to the model, there are two reasons why children have reading comprehension difficulties. First, children may

have difficulty accessing meaning from written material, which is secondary to a decoding deficit and second, they may have weaknesses in linguistic comprehension. It is well established that inference generation is a critical component of linguistic comprehension. It follows that this model predicts how children will perform on inference tasks based on their reading skill profile.

In poor comprehenders, previous research clearly demonstrates that they have difficulties generating inferences when reading so one would expect to see performance differences in this group. More specifically if poor comprehenders have difficulties making inferences as they read (e.g. Oakhill, 1982, 1983, 1984) and do not monitor their comprehension effectively (e.g. Oakhill, Hartt & Samols, 2005; Baker & Brown, 1984), they would be expected to show less inhibition of reading when they encounter an inconsistency in the text. This would be due to them either failing to detect it as they have failed to make the inference required for the inconsistency to be apparent, or because they ignore it rather than attempting to resolve it.

In poor decoders, there is very little published evidence that specifically considers their ability to generate inferences, and none suggests that inferential skills are affected when decoding is poor. Furthermore, no previous published studies were found that investigated inference generation in poor decoders using an on-line measure. Catts, Adlof and Ellis Weismer (2006) reported that poor decoders' performance in discourse comprehension and inferencing tasks was equivalent to that of typical readers. Nation and Snowling (1998b) found that poor decoders showed increased contextual facilitation of word recognition than either typical readers or poor comprehenders. Both of these studies are consistent with the notion that poor decoders' inferential processing is unaffected by their reading difficulty. Thus, providing decoding is successful, poor decoders and typical readers would be expected

to demonstrate equivalent levels of inferential activity, indexed by similar levels of inhibition of reading when they encounter inconsistency.

To test these hypotheses, a self-paced reading paradigm was used. Children read or heard short stories presented sentence by sentence and were required to press a key to advance to the next sentence. Reading/listening times to target sentences were recorded. Target sentences were either consistent with passage context or introduced an inconsistency. Inconsistent target sentences were very closely matched to consistent target sentences, to allow measurement of the amount of interference induced by inconsistency. The differences between time to read inconsistent and consistent target sentences were compared between groups to test whether reading skill profile had any impact on inferential processing. Crucially, if an inconsistency slows a reader/listener down, this demonstrates that the relevant inference must have been generated during reading of the prior text, and it can be inferred that they are making an attempt to resolve it. Previous work using this paradigm conducted by Gernsbacher and colleagues was used to successfully establish that adults make emotion-state inferences on-line during reading (Gernsbacher, Goldsmith & Robertson, 1992; Gernsbacher, Hallada & Robertson, 1998).

A difficulty in making a comparison between the processing of text by poor decoders and poor comprehenders is constructing materials that are accessible to poor decoders. Throughout this thesis the difficulty levels of any texts that children were asked to read were carefully controlled to ensure that all participants (including poor decoders) were capable of decoding them and accessing their content. It has been shown that in the very early stages of learning to read, greater resources are allocated to decoding, and persistent difficulties with comprehension often only become evident once decoding is reasonably competent, a phenomenon described by Leach,

Scarborough and Rescorla (2003) as the ‘fourth grade slump’. The age range selected for these studies was chosen so that difficulties with comprehension would be measurable and clear, and any difficulties with decoding could be classed as persistent or unusual rather than reflecting a typical developmental trajectory.

#### *6.1.1 Overview of Findings*

Experiments 1 and 2 assessed whether children could generate inferences about fictional characters’ emotion states and spatial inferences on-line while reading short narratives, and investigated the influence of decoding and comprehension skills on task performance. Reading times were compared to carefully controlled target sentences that were either consistent or inconsistent with previous passage context. Crucially, any inconsistency would only be noticed if the reader generated inferences about the protagonist’s emotion-state or spatial inferences on-line.

In line with prediction, typical readers read target sentences more slowly if they were inconsistent with context. Moreover, although poor decoders read all of the passages more slowly than the typical readers, their reading was disrupted to a similar extent by the presence of inconsistent material. In contrast, poor comprehenders were less inhibited by inconsistency. It could be argued that this was an artefact of their faster reading rate, but the differences remained even when reading speed was accounted for. The results suggest that poor comprehenders can generate inferences on-line, but that they either do this less frequently, or make less effort to resolve any inconsistency with those on-line inferences. Thus, Experiment 1 established the validity of the paradigm for investigating inference generation in this population, and demonstrated that typical readers and poor decoders generated emotion-state inferences on-line. Poor comprehenders had difficulty generating emotion-state inferences on-line, consistent with their comprehension weaknesses.

These findings were replicated in Experiment 2 using spatial inferences. Typical readers and poor decoders showed similar levels of inhibition of reading when target sentences were inconsistent, whereas poor comprehenders showed significantly less inhibition. This confirmed that the poor comprehenders' difficulties were not specific to emotion-state inferences.

Experiment 3 investigated whether knowledge-based inferences were generated on line, and whether timing of acquisition of the background knowledge affected the ability to generate such inferences on-line. Age of acquisition was manipulated by comparing inferences about general knowledge that children would be expected to know on school entry, and general knowledge that was taught in the first term of Year 4. Contrary to the prediction that only poor comprehenders would have difficulties inferring from recently acquired general knowledge, all groups found it difficult to generate inferences on-line if the general knowledge required to do so was recently acquired, and showed no inhibition of reading to inconsistencies based on recently acquired general knowledge.

Experiment 4 sought to replicate this finding in a different sample of children with the poor decoder, typical reader and poor comprehender profiles. In addition, since all groups tested in Experiment 3 failed to show inhibition when reading inconsistent target sentences in the recently acquired condition, Experiment 4 sought to determine whether any response to the inconsistency was delayed rather than absent, by controlling the content of the post-target sentences and examining their reading times.

In the poor comprehenders, the findings of Experiment 3 were replicated; poor comprehenders showed inhibition when reading inconsistent target sentences in the early-acquired condition, but not in the recently acquired knowledge condition.

However, in contrast to the findings of Experiment 3, the typical readers and the poor decoders generated inferences in both the early acquired and recently acquired conditions. The findings of Experiment 4 were in line with original prediction; only the poor comprehenders failed to show a consistency effect in the late-acquired condition, whereas the other groups showed a robust consistency effect in both conditions.

Examination of reading times and inhibition scores to post-target sentences showed that there was greater inhibition of reading in the presence of inconsistency in the late-acquired condition than in the early-acquired condition, suggesting that the integration of more recently learned information with incoming text is slower therefore carry-over can be seen in the post-target sentences. However poor comprehenders do not simply show slower processing indexed by slower reading of post-target sentences, they seem to make less overall effort to resolve inconsistencies than either poor decoders or typically developing readers.

A factor that could account for the discrepancy between Experiments 3 and 4 was that in Experiment 4 the knowledge base required for integration of the later acquired information had been more recently reinforced in classroom work, as these children were tested earlier in the school year. The ability to use general knowledge to make inferences could therefore depend upon an interaction between the length of time that an individual has had to integrate that knowledge and how recently and frequently it has been reinforced.

Experiment 5 was conducted to determine whether the findings in the written modality could be replicated in the oral modality. The results of Experiment 5 showed that the differences in performance shown by the poor comprehenders are not confined to the written modality, and similar difficulties with inference generation are evident

in the oral modality as well. This is in line with the Simple View of Reading, which implies that comprehension difficulties permeate all linguistic modalities.

An unexpected finding was that the poor decoders had greater difficulty with the auditory task, and were not able to generate inferences on-line when the linguistic material was presented orally. This was inconsistent with their performance on two off-line measures of comprehension. First, in a background measure of listening comprehension, the poor decoders scored as highly as typical readers, and second they were as good as the typical readers overall at answering the comprehension questions posed after each story. However, finer grained analysis of their performance on the comprehension questions revealed a subtle weakness in their ability to generate coherence inferences off-line. This fits with published evidence that suggests that poor decoders have subtle underlying language weaknesses in addition to their decoding difficulties. An alternative view is that co-morbid attention difficulties might be responsible for the lack of on-line inference generation in the oral task. It is suggested that the reasons for the discrepant performance of poor decoders in the on-line oral inference generation task merits further investigation. For example, it might be useful to make a comparison with groups matched on verbal IQ to exclude the possibility that small, (but nonsignificant) differences in verbal IQ between the typical readers and the poor decoders could account for the differences observed in Experiment 5.

Together, the results of Experiments 1-5 are consistent with previous published research on inference generation in poor comprehenders in that poor comprehenders engaged in less inferential activity than typical readers (e.g. Yuill & Oakhill, 1991; Cain & Oakhill, 1998, 1999; Cain, Oakhill, Barnes & Bryant, 2001). The findings are also consistent with research on comprehension monitoring, suggesting that poor comprehenders do not monitor their comprehension effectively (e.g. Oakhill, Hartt &

Samols, 2005; Cain, Oakhill & Bryant, 2004; Hacker, 1997; Yuill, Oakhill & Parkin, 1989; Garner & Kraus, 1981).

A compelling hypothesis that would largely explain the performance differences in poor comprehenders in Experiments 1-5 is that they apply a lower standard for coherence, and therefore do not monitor their comprehension well or generate as many inferences as good comprehenders. Experiment 6 was carried out to test this idea. Children read short vignettes that were followed by a question. They were then asked to evaluate each of four possible answers to each question that varied in plausibility, by either indicating 'yes' if they felt they were in any way plausible, or 'no' if they were implausible. The valence and response times to perform this task were compared. The faster reaction times shown by poor comprehenders compared with typical readers and poor decoders lends support to the hypothesis that they do apply a reduced standard for coherence when comprehending text. In Experiment 6, in conditions 2 and 3 where the correct response was uncertain, they spent less time deliberating about the response than the other two groups. Poor comprehenders seemed to find it easier to deal with uncertainty about what constituted a correct answer and their faster response times indicate that they arrive at their response after less consideration of the implications of each response.

Interestingly in this experiment, the valence of poor comprehenders' responses in all plausibility conditions were generally similar to those of the other groups, so in effect they performed better on this task than the other groups (i.e. equivalent accuracy was achieved more quickly). However, this task only measured responses and reaction times to possible answers to questions following very short vignettes that were easy for all participants to decode (mean length = 24 words), and therefore this task did not require participants to construct elaborate situation models to fully appreciate each



scenario. It may be that the accuracy of poor comprehenders' responses would differ from that of typical readers or poor decoders if the passages were more difficult or longer if they applied a similar standard for coherence. Furthermore, the explicit nature of the task may have facilitated poor comprehenders' performance, and it is possible that may not generalize to text reading tasks where task demands are more implicit.

It was hoped that further information about the performance of these children could be gleaned by examining patterns of correlations between participant characteristics measured using standardised tests and also between them and the outcome measures from Experiments 1-6. A meta-analysis of participant characteristics was also conducted but no coherent patterns of correlations were found. This may be due to the size of the samples taking part in these studies, as larger scale intervention studies have found coherent patterns of correlations that enable testable hypotheses to be generated for further study.

#### *6.1.2 Theoretical Context*

The results presented in this thesis lend support for the constructionist position (e.g. Graesser, Singer & Trabasso, 1994) by showing that developing readers strive for local and global coherence, and make a wide range of on-line inferences automatically during reading. The findings reported are consistent with those reported by O'Brien, Albrecht and colleagues (O'Brien & Albrecht, 1992; Albrecht & O'Brien, 1993; Myers, O'Brien, Albrecht & Mason, 1994; Albrecht & Myers, 1995) using a similar paradigm in adult readers, as children were found to be sensitive to violations in global coherence. The minimalist view (e.g. McKoon & Ratcliff, 1992) proposes that elaborative and coherence inferences that are not required for local coherence will not be generated on-line, therefore the findings reported here do not support this view.

However, it is important to point out that there are a range of circumstances under which minimalist theory predicts that these inferences could be made on-line, which makes it extremely difficult to rule this theoretical position out.

The results from poor comprehenders show that they generate fewer inferences on-line, and previous research manipulating the distance between premises requiring integration (e.g. Cain & Oakhill, 1999) show that they have difficulties when premises are further apart in the text. It is suggested that their reading behaviour appears more consistent with the minimalist view, however there is still evidence that they generate inferences on-line beyond those that would be predicted by the minimalist hypothesis.

### *6.2 Differences Between Off-line and On-line Processes*

In Experiments 1-5, a measure of off-line comprehension was taken, which established that participants were reading for meaning, or listening attentively. Four questions were asked after each story, two requiring inferences (one coherence, one elaborative), and two requiring participants to supply factual information. All groups found the inferential questions more difficult. For each experiment poor comprehenders answered fewer questions correctly than either the poor decoders or the typical readers, who were equally accurate at answering the questions. This provides support for the selection protocol applied, and is consistent with the results of the on-line inferencing task, showing that poor comprehenders were worse at generating inferences than the other two groups.

In the oral modality, consistent with their results in the on-line inferencing task, poor comprehenders answered fewer comprehension questions correctly than the typical readers and the poor decoders, who performed at a similar level. However, poor decoders had greater difficulty making inferences on-line when listening to oral language, but their off-line comprehension appeared intact. Although poor decoders'

overall accuracy when answering comprehension questions was similar to that of typical readers, they were poorer at answering coherence inference questions.

### *6.2.1 Individual Differences and the Development of Comprehension Difficulties*

The research outlined in Chapter 1 of this thesis demonstrates that linguistic comprehension requires the integration of a range of basic and higher-level component language skills. It is plausible therefore that a deficit in any of these component skills may lead to comprehension deficits. According to Perfetti, Landi and Oakhill (2007), groups of readers classified as ‘poor comprehenders’ contain children with a wide variety of skill profiles (Cain & Oakhill, 2006a; Nation, Clarke, Marshall & Durand, 2004; Cornoldi, de Beni & Pazzaglia, 1996). This must be borne in mind when applying generalisations about poor comprehenders at the individual level. Poor comprehenders identified on the basis of their performance in reading comprehension tests often share some common underlying problems such as difficulties generating inferences and weaknesses in comprehension monitoring, in the presence of intact phonological skills (e.g. Catts, Adlof & Ellis Weismer, 2006; Cain, Oakhill & Bryant, 2000; Stothard & Hulme, 1995).

#### *6.2.1.2 Individual differences, inferencing difficulties and comprehension*

In the studies presented in this thesis, poor comprehenders consistently seem to make fewer inferences on- and off-line than both typical readers and poor decoders. Furthermore, the data are consistent with the view that poor comprehenders do have the prerequisite skills to make inferences but show less propensity to do so; their performance shows a reduced rather than an absent effect of consistency, and their performance on the off-line inference questions indicates a reasonable level of success but reduced performance compared with the typical readers and poor decoders. There are a number of factors that could plausibly lead to this pattern of performance.

Effortful decoding could result in fewer resources being available for additional processing such as the generation of inferences. However, both in the studies reported here and in the wider literature poor comprehenders do not show any evidence of decoding or phonological difficulty. There is also no consistent evidence in longitudinal studies of any phonological difficulties occurring and resolving prior to the development of reading comprehension deficits (e.g. Nation, Cocksey, Taylor & Bishop, 2010 in press). In the experiments reported in this thesis, poor comprehenders decoding skills were at least in line with those of typically developing readers. There was no evidence that their reading was any more laboured, and there was a tendency for poor comprehenders to decode faster than typical readers.

The children's level of background knowledge is likely to affect their ability to generate knowledge-based coherence and elaborative inferences. However studies where the knowledge base has been tightly controlled and taught to criterion have demonstrated that even when it has been shown empirically that poor comprehenders have equivalent absolute levels of background knowledge, they still perform poorly on inferencing tasks (e.g. Cain, Oakhill, Barnes & Bryant, 2001; Barnes, Dennis & Haefaele-Kalvaitis, 1996). However, as Hulme and Snowling (2009) point out, this does not mean that the knowledge is equally accessible to them, given that the way in which they store the knowledge in memory cannot be controlled experimentally.

The role of vocabulary in reading comprehension is clearly an important one; if prerequisite vocabulary is lacking, understanding a passage of text is likely to be impaired, with knock-on effects on the ability to generate inferences about its content. A further interaction between vocabulary and inferencing is shown by studies of new word learning, which demonstrate that poor comprehenders have difficulty in inferring the meanings of new words from context (e.g. Nation, Clarke & Snowling, 2007;

Nash & Snowling, 2006; Cain, Oakhill & Elbro, 2003). Thus, the relationship between vocabulary and reading comprehension is mediated by inferential skills, and the ability to generate inferences is moderated by knowledge of the world. In both samples of poor comprehenders taking part in the studies presented in this thesis, vocabulary was lower than in the typically developing readers, although this difference was not consistently statistically significant. However it is unlikely that poor vocabulary was responsible for the differences in inferencing performance in these studies, as the vocabulary contained within the stories presented was simpler than would be used in a typical age-appropriate text, in order that poor decoders were capable of reading them.

Given that no consistently occurring underlying cognitive factors have been identified that predict *reduced* inference generation (as opposed to an inability to generate particular inferences) in poor comprehenders, the results from experiment 6 are particularly pertinent. Experiment 6 explores the possibility that reduced levels of inference generation might reflect the application of a lower standard for coherence by poor comprehenders during reading. The results from this study suggested that poor comprehenders spent less time evaluating their response to test items where there was an element of uncertainty, which would be consistent with this view. Crucially evaluation of the test items required participants to make inferences based on the content of short passages, and therefore the results suggest that the poor comprehenders engaged in less of this inferential activity.

If one were to accept the view that a reduced standard for coherence led to a reduction in inferential processing, it would be necessary to explore the reasons why a child might learn to accept a reduced standard for coherence when processing linguistic input. First, there may be cognitive factors that lead them to become

accustomed to a paucity of meaning in the language input that they are exposed to; oral language weaknesses in early development have been demonstrated in longitudinal studies following children that subsequently develop reading comprehension difficulties (Nation et al, 2010 in press; Catts et al, 2006), for example. Second, environmental factors such as motivation to read for meaning and levels of exposure to language and print could also mediate the standard for coherence applied. It is not clear whether motivation is a factor for poor comprehenders, however the demonstrated oral language weaknesses may effectively reduce exposure to coherent language input and also reduce a child's desire to read, given that their experience of language is less rich and meaningful than it should be. It is suggested that it would be useful to explore these factors in more detail in future studies.

In summary, from the wider literature and the studies presented in this thesis, a clear picture of the causes of inferencing difficulties reported in poor comprehenders is not forthcoming. However, it is suggested that reciprocal interaction between inferential processing, comprehension and the standard for coherence being applied could plausibly reduce the drive to generate inferences automatically during reading. The results of the experiments presented in this thesis would be consistent with this view.

### *6.2.2 Decoding Difficulties*

For decoding difficulties, the picture is somewhat clearer. Extensive published research has confirmed that a phonological deficit lies at the root of poor decoding (see Hulme & Snowling 2009 for review). However, there is very little published research that focuses on the comprehension skills of poor decoders specifically. In this thesis poor decoders on-line inferencing skills were found to be equivalent to that of typical readers in the written modality. This fits well with theories of dyslexia that

posit poor phonology in the presence of intact language processing. The results of Experiment 5 in the auditory modality suggest that further research of poor decoders' oral language comprehension is warranted, as at least for some, impairments in the auditory modality may have been overlooked. This may reflect heterogeneity in the poor decoder sample, which may have included children with dyslexia with co-morbid language impairments, or generally poor readers. According to the classification of language disorder proposed by Bishop and Snowling (2004), these children would be expected to show nonphonological language difficulties in addition to their reading difficulties.

### *6.3 Implications*

The studies reported here show that poor comprehenders have difficulties generating inferences on-line when they read or listen to short narratives. The results of Experiments 1-5 are consistent with research evidence showing that poor comprehenders can generate inferences, however their inferencing is less robust and they do not generate the same number or range of inferences effortlessly and automatically that typical readers seem to do. This difficulty is clearly not specific to the written domain, as in Experiment 5 poor comprehenders were shown to generate fewer on-line inferences when listening to short narratives. The comprehension difficulty observed here does not, therefore, reflect an underlying decoding issue such as disproportionate resources being required for adequate word recognition (Perfetti, 1985).

If one considers inference generation to be an essential part of coherence building, then the number of inferences generated will depend upon the standard for coherence adopted by the reader or listener (Van den Broek, Risdén & Husebye-Hartmann, 1995). According to this view, the standard for coherence adopted for the

reading or listening task therefore drives the number of inferences and potentially the type of inferences that are generated on-line. This view, consistent with data reported here, would suggest that inference generation is an implicit activity rather than a consequence of metacognition, whereas the setting of parameters based on the standard for coherence adopted for a given language-based task forms the metacognitive component. Similarly, comprehension-monitoring activity would be dictated by the standard for coherence adopted.

The results of Experiment 6 provided some support for the view that poor comprehenders do set a lower standard for coherence when reading. It follows that they do not engage in as much activity designed to integrate current and preceding text than typical readers, and are less concerned by inconsistencies present in the text. They are more accepting of inconsistencies and less likely to make concerted efforts to resolve them.

The reasons for children to develop the habit of adopting a lower standard for coherence are likely to vary, however, it would seem reasonable to suggest that interventions targeting use of reading strategies that would elevate standard for coherence might be useful in boosting their overall comprehension levels. If the argument holds that it is an individual's standard for coherence that drives them to generate inferences, then such intervention should be more successful than inference training, which has been demonstrated to boost comprehension in less skilled readers (e.g. Yuill & Oakhill, 1998; McGee & Johnson, 2003). To date, there are no published interventions targeting standard for coherence specifically, however emerging findings from interventions targeting reading strategies such as iSTART (Interactive Strategy Training for Active Reading and Thinking - McNamara, Levinstein & Boonthum, 2004) have shown promise. For example, Taylor, O'Reilly,



Rowe and McNamara (2009) found that using iSTART to train use of five key reading strategies (comprehension monitoring, paraphrasing, prediction, bridging inferences, and elaboration) led to improvements in expository text comprehension of adolescent readers with low domain knowledge (Taylor, O'Reilly, Rowe & McNamara, 2009). Similarly, Clarke, Hulme, Truelove & Snowling (2009) compared the efficacy of three interventions to remediate comprehension difficulty. Their text comprehension programme comprised activities taught using the framework of Palincsar and Brown's (1984) reciprocal teaching approach (clarifying, summarising, predicting, and question generation) and included four key taught elements; inferencing from text, reading comprehension, written narrative and explicit teaching of metacognitive strategies. They found that participants made significant and sustained gains in comprehension skill compared with waiting controls.

If comprehension can be improved by teaching strategies, it is plausible that poor comprehenders possess the prerequisite skills for good comprehension, but lack awareness of which skills need to be engaged, and when to engage them. This is consistent with Cataldo and Oakhill's (2000) assertion that poor comprehenders do not match their reading behaviours to their reading goals in the same way that good comprehenders do. Lorch, Klusewitz and Lorch (1993) identified 10 classes of reading situation, and found that readers varied the standard for coherence adopted according to task demands (Lorch, Klusewitz & Lorch, 1995). It follows that if readers do not have either the knowledge or the ability to adopt a task-appropriate standard for coherence then there will be shortfalls in their comprehension.

When learning to read, children need to develop a full understanding of what reading can give to them in terms of meaning. For some, this develops naturally as they engage with the reading process. For others, this may need to be explicitly

taught, by encouraging children to generate inferences, monitor comprehension and explore what additional information they can get from a story by integrating their general knowledge with language input. In short, some children need to be taught what it means to comprehend, and to fully appreciate what they can gain from being able to read well. If they do not acquire this insight at an early age, then their skills gap will widen due to Matthew effects (Stanovich, 1986). It is suggested that a program to encourage children to adopt a high standard for coherence could be implemented on a whole class basis during early reading instruction in a bid to thwart the development of comprehension difficulties in vulnerable individuals.

If a low standard for coherence is at the root of comprehension difficulties, then considering how known developmental precursors to comprehension difficulties might impact upon a child's standard for coherence might yield worthwhile avenues for further research.

#### *6.4 Limitations*

The studies reported here compare chronological age-matched groups. A more rigorous methodology would also include comparison groups matched on the skill level of the disordered groups in order to exclude lack of experience as a potential explanation for any differences observed (Cain, Oakhill & Bryant, 2000; Bradley & Bryant, 1978). However, given the time constraints and practicalities involved in recruiting 5 groups for each experiment it was decided to focus on comparisons of 2 reading disordered groups with chronologically age-matched controls.

The experimental materials had to be designed to be easy enough for poor decoders to read in order to allow the three groups to be compared therefore task difficulty was constrained. The materials would have been very easy for the poor comprehenders and typical readers to decode, and less demanding than texts they

commonly read in school. It also means that the language used was relatively simple both in terms of vocabulary and syntax. This does however make it less likely that the differences seen in the poor comprehenders are the result of underlying decoding problems.

It was possible that the poor decoder groups tested contained both garden-variety poor readers and poor decoders. Children were selected based on their performance on the Neale Analysis of Reading Ability, and had to show a discrepancy between actual achievement and that which would be predicted for their age. The view was taken that their reading comprehension would be constrained by their reading accuracy, therefore comprehension in line with reading accuracy was accepted. This means that children with broader based reading difficulties may have been included. It is suggested that in future experiments selection should be based on poor decoding in the presence of normal listening comprehension. Examination of the data from individual participants suggests that only three of the children taking part in Experiments 4-6 showed below average levels of listening comprehension so it is unlikely that this had a significant impact on the findings. A background measure of listening comprehension was not obtained for Experiments 1-3 therefore it is not possible to examine this further for poor decoders in these experiments.

Finally, a limitation of passage reading tasks such as those reported here is that in order that the passages are naturalistic and read like 'normal' stories, there has to be compromise on what elements of the passages to control. In the studies reported here, the content of the test sentences, passage length, and word length were very tightly controlled, however factors such as frequency and age of acquisition of all vocabulary used, and syntactic structure were not. It is possible that differences between the passages may have had some effect on inference generation; for example word-based

priming may have varied due to the different context in which target sentences were embedded. However, the present results are consistent across all experiments reported and with experiments using similar methodology in the published literature.

### *6.5 Suggestions for Further Research*

The experiments using reading time provide strong evidence that poor comprehenders engage in less inferential processing than either poor decoders or typical readers. However, a limitation of this methodology is that although it can be demonstrated that inferencing activity leads to inhibition of reading in typical readers, it is not possible to determine what additional activity is being engaged in when reading is slowed. Experiments comparing poor comprehenders and typical readers using this paradigm in combination with eye-tracking may reveal more about inferential activity in both groups, as eye movements will reveal which areas of the text readers are focussing on.

The reading time experiments could be made more robust by adding more stories, and by the inclusion of additional neutral or filler material. Gernsbacher and Robertson (1992) showed, in typical readers, the disruption seen when an inconsistency is presented is greater when the proportion of stories containing inconsistency is lower, so results generated using more filler material may yield additional information. This would be possible if similar experiments were run with more competent decoders, but was not possible here because materials had to be accessible and not too arduous for the poor decoders. Essentially, there had to be compromise on quantity and difficulty of material so that poor decoders could be included thus providing a test of predictions made by the Simple View of Reading.

It would be useful to conduct further research into oral language comprehension skill in poor decoders. Contrary to established belief, the results of

Experiment 5 suggested that oral language comprehension is compromised in poor decoders, and that this may be due to oral language weaknesses that are not tapped in the written modality. This may reflect the fact that poor decoders may be able to control the speed at which they receive and process language input when they are decoding it for themselves whereas they cannot slow the pace when they are listening to spoken language. Poor decoders may have weaknesses affecting speed of language processing, which are not picked up by conventional reading assessments.

It is possible that poor comprehension skills could reflect a trade-off between reading accuracy and reading comprehension, which has a greater impact on comprehension as reading skills develop, and task demands increase. Children who can decode well receive a lot of positive reinforcement from their teachers during early reading instruction, which may lead them to focus more on their decoding skills, to the detriment of reading for meaning. The performance of the poor comprehenders taking part in Experiments 1-3 on the TOWRE lends support for this view, as despite being matched on an untimed passage reading accuracy test, their sight word reading and nonword decoding was significantly faster than that of the typical readers. The participants in Experiments 4-6 were, on average, 14 months younger, and did not show this difference, which suggests that if this argument holds, then this difference emerges later. It would be useful to investigate this further, as if it was found to be the case that positive reinforcement was leading to a disproportionate allocation of resources to the development of decoding skills in these children, then early reading instruction would need to be modified to place greater emphasis on positively reinforcing the value of reading for meaning, and training of metacognitive strategies.

Further research is needed to clarify the standard for coherence adopted by poor comprehenders across different text genres and texts of different lengths and difficulty

levels, in order to establish whether they operate a reduced standard across the range of written materials. In Experiment 6, poor comprehenders showed a reduced standard for coherence, but their absolute accuracy on the task was no different to that of poor decoders and typical readers. Arguably, adopting a reduced standard for coherence did not have deleterious effects on this task because of the simplicity and brevity of the vignettes or the explicit nature of the task. It would be interesting to compare how poor comprehenders modify their standard for coherence according to implicit task demands and compare that with typical readers.

### *6.6 Conclusions*

This thesis reports a series of experiments that consistently show that poor comprehenders make fewer inferences on-line when reading short narratives than either poor decoders or typically developing children. They also make fewer inferences on-line than typically developing children when listening to short narratives. One factor that influences their propensity to generate inferences is the recency of knowledge acquisition. All groups found it more difficult to make inferences requiring integration of recently acquired knowledge, but this was more pronounced in poor comprehenders.

This thesis has also extended previous work by applying the methodology used by Gernsbacher and colleagues (e.g. Gernsbacher, Goldsmith & Robertson, 1992) to measure on-line inference generation in developing readers, and by using it to assess on-line inference generation in reading-disabled groups. Furthermore, the design used in all of these studies, comparing two reading-disabled groups with typical readers, has only rarely been used, and this is the first time that on-line inference generation has been investigated in both poor decoders and poor comprehenders together.

No previous published research has been found which examines the on-line inferencing skills of poor decoders. The results presented here show for the first time that poor decoders do make inferences on-line during reading like typical readers. However, subtle oral language weaknesses may compromise their performance when required to generate inferences during listening, and when task demands are high.

It has been argued that poor comprehenders' performance is due to them adopting and implementing a lower standard for coherence than poor decoders and typically developing children. As van den Broek, Risdén and Husebye-Hartmann (1995) suggest, this would logically lead to a reduced propensity to generate inferences on-line. The results of all the experiments reported in this thesis are consistent with this view.

To summarise, the research presented here adds to the evidence for the validity and utility of classifying reading disability according to the framework suggested by the Simple View of Reading. The experiments reported in this thesis provide a clear demonstration that children with decoding difficulties show no impairment in inference generation during reading whereas poor comprehenders show significant difficulties generating a range of inferences both on-line and off-line. The results are consistent with the view that poor comprehenders generate fewer inferences on-line than poor decoders and typical readers because they apply a lower standard for coherence when processing spoken and written language.

## References

- Adlof, S M., Catts, H.W., Hogan, T.P., & Little, D.T. (2005, June) *The role of fluency in reading comprehension: Should fluency be included in the simple view?* Poster presented at the annual meeting of the Society for the Scientific Study of Reading, Toronto, ON, Canada.
- Albrecht, J.E., & O'Brien, E.J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 1061-1070.
- Albrecht, J.E. & Myers, J.L. (1995). The role of context in the accessibility of goals during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1459-1468.
- Baker, L., & Brown, A.L. (1984). Metacognitive skills and reading. In P.D. Pearson, M.L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research*. White Plains, NY: Longman.
- Barnes, M.A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferences in children from six to fifteen years of age. *Journal of Experimental Child Psychology*, 61, 216-241.
- Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74, 506-521.
- Bishop, D.V.M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry*, 31, 1027-1050.



- Bishop, D.V.M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, 52, 156-173.
- Bishop, D.V.M., & Snowling, M.J. (2004) Developmental Dyslexia and Specific Language Impairment: Same or Different. *Psychological Bulletin*, 130, 858-886.
- Bowyer-Crane, C. A. (2002). The relationship between reading comprehension and on-line inference generation in children. Unpublished Doctoral Thesis, The University of York, 2002.
- Bradley, L., & Bryant, P.E. (1978). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature*, 271, 746-747.
- Bransford, J.D., & Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behaviour*, 11, 717-726.
- Bransford, J.D., Barclay, J.R., & Franks, J.J. (1972). Sentence memory: A constructive versus interpretative approach. *Cognitive Psychology*, 3, 193-209.
- Buzan, A. (1974) *Use Your Head*. London, UK: BBC Publishing.
- Caccamise, D., & Snyder, L. (2005). Constructive models of reading comprehension – the situation model and related constructions. *Topics in Language Disorders*, 25, 1.
- Cain, K. (1994). *An investigation into comprehension difficulties in young children*. Sussex, UK. Unpublished PhD dissertation, University of Sussex.
- Cain, K. (1996). Story knowledge and comprehension skill. In C. Cornoldi, & J.V. Oakhill (eds), *Reading Comprehension Difficulties: Processes and Interventions*. Mahwah, NJ: Erlbaum.
- Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology*, 21, 335-351.

- Cain, K., & Oakhill, J. (1996). The nature of the relationship between comprehension skill and the ability to tell a story. *British Journal of Developmental Psychology*, 14, 187-201.
- Cain, K., & Oakhill, J.V. (1998). Comprehension Skill & Inference-Making Ability. In Hulme, C., & Joshi, R.M., (eds) *Reading and Spelling: Development and Disorders*. London. Lawrence Erlbaum Associates.
- Cain, K. and Oakhill, J. (1999). 'Inference making ability and it's relation to comprehension failure in young children', *Reading and Writing: An Interdisciplinary Journal*, 11, 5-6, 489-503.
- Cain, K., & Oakhill, J. (2006a) Profiles of children with specific reading comprehension difficulties, *British Journal of Educational Psychology*. 76, 683-696.
- Cain, K., & Oakhill, J. (2006b) Assessment Matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*. 76, 697-708.
- Cain, K., Oakhill, J.V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, 29, 850-859.
- Cain, K., Oakhill, J., & Bryant, P. (2000). Phonological skills and comprehension failure: a test of the phonological processing deficit hypothesis. *Reading and Writing*, 13, 31-56.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's Reading Comprehension Ability: Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology*, 96(1), 31-42.
- Cain, K., Oakhill, J.V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language*, 30. 681-694.
- Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of

- word meanings from context: the influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, 96, 671-681.
- Casteel, M.A. (1993). Effects of inference necessity and reading goal on children's inferential generation. *Developmental Psychology*, 29, 346-357.
- Cataldo, M.G. and Oakhill, J. (2000). Are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology*, 92, 791-799.
- Catts, H., Adlof, S., & Ellis Weismer, S. (2006). Language deficits in poor comprehenders: a case of for the simple view. *Journal of Speech, Language & Hearing Research*, 49, 278-293.
- Catts, H.W., Hogan T.P., & Adlof, S.M. (2005). Developmental Changes in Reading and Reading Disabilities. In Catts, H.V & Kamhi (Eds). *The Connections between language and reading disabilities*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Catts, H., Hogan, T.P., & Fey, M. (2003). Subgrouping poor readers on the basis of reading-related abilities. *Journal of Learning Disabilities*, 36, 151-164.
- Catts, H., & Kamhi, A. (2005). (Eds.) *The connections between language and reading disabilities*. Mahway, NJ: Lawrence Erlbaum Associates.
- Chall, J.S. (1983). *Stages of Reading Development*. New York. McGraw-Hill.
- Clarke, P.J., Hulme, C., Truelove, E. & Snowling, M.J. (2010 in press). Efficacy of Interventions to Promote Reading Comprehension: A Randomised Controlled Trial. *Psychological Science*.
- Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In Cornoldi, C., & Oakhill, J.V. (Eds.),

*Reading comprehension difficulties: Processes and interventions* (pp. 113-136).

Mahwah, NJ: Erlbaum.

Cragg, L., & Nation, K. (2006) Exploring written narrative in children with poor reading comprehension, *Educational Psychology*, 26, 55-72.

Cunningham, A.E., & Stanovich, K.E. (1998) The impact of print exposure on word recognition. In Metsala, J.L., & Ehri, L.C. (eds.) *Word recognition in beginning literacy*. Mahwah, NJ: Lawrence Erlbaum Associates.

Curtis, M.E. (1980). Development of components of reading skill. *Journal of Educational Psychology*, 72, 656-669.

Daneman, M. (1988). Word knowledge and reading skill. In Daneman, M., MacKinnon, G. & Waller, T.G. (Eds.), *Reading research: Advances in theory and practice*. (Vol. 6, pp. 145-175). San Diego: Academic Press.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, 19, 450-466.

De Jong, P. F. & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, 6, 51-77.

de Sousa, I., & Oakhill, J.V. (1996). Do levels of interest have an effect on children's comprehension monitoring performance? *British Journal of Educational Psychology*, 66, 471-482.

Durand, M., Hulme, C., Larkin, R., & Snowling, M.J. (2005). The cognitive foundation of reading and arithmetic skills. *Journal of Experimental Child Psychology*, 91, 113-136.

- Freebody, P., & Anderson, R.C. (1983). Effects of vocabulary difficulty, text cohesion and schema availability on reading comprehension. *Reading Research Quarterly*, 14, 624-645.
- Frijda, N. (1986). *The Emotions*. Cambridge, UK: Cambridge University Press.
- Garner, R., & Kraus, K. (1981). Good and poor comprehender differences in knowing and regulating reading behaviors. *Education Research Quarterly*, 6, 5-12.
- Gayán, J., & Olson, R. K. (2003). Genetic and environmental influences on individual differences in printed word recognition. *Journal of Experimental Child Psychology*, 84, 97-123.
- Gernsbacher, M.A. (1990) *Language comprehension as structure building*. Hillsdale, NJ. Lawrence Erlbaum Associates.
- Gernsbacher, M.A. (1995) "Activating Knowledge of Fictional Characters' Emotional States." In Weaver, C.A., Mannes, S., & Fletcher, C.R. (eds) (1995) "Discourse Comprehension: Essays in Honor of Walter Kintsch." Hillsdale, Lawrence Erlbaum Associates.
- Gernsbacher, M.A. & Faust, M.E. (1991). The mechanism of suppression: a component of general comprehension skill. *Journal of Experimental Psychology: Learning Memory and Cognition*. 17(2), 245-262.
- Gernsbacher, M.A., Goldsmith, H.H., & Robertson, R.R.W. (1992). Do Readers Mentally Represent Characters' Emotional States? *Cognition and Emotion*, 6(2), 89-111.
- Gernsbacher, M.A., Hallada, B.M., & Robertson, R.R.W. (1998). How Automatically Do Readers Infer Fictional Characters' Emotional States? *Scientific Studies of Reading*, 2(3), 271-300.

- Gernsbacher, M.A., & Robertson, R.R.W. (1992). Knowledge Activation versus Sentence Mapping when Representing Fictional Characters' Emotional States. *Language and Cognitive Processes*, 7, 353-371.
- Gilger, J.W., Pennington, B.F., & DeFries, J.C. (1992) A twin study of the etiology of comorbidity: Attention deficit hyperactivity disorder and dyslexia. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 343-349.
- Goff, D.A., Pratt, C., & Ong, B. (2005). The Relations Between Children's Reading Comprehension, Working Memory, Language Skills and Components of Reading Decoding in a Normal Sample. *Reading and Writing*, 18, 7-9, 583-616.
- Gordon, C.J. & Braun, C. (1985). Metacognitive processes: Reading and writing narrative discourse. In D.L. Forrest-Pressley, G.E. MacKinnon, & G. Waller (Eds.), *Metacognition, cognition and human performance*. New York: Academic Press.
- Gough, P.B., Hoover, W.A., & Peterson, C.L. (1996). Some observations on the simple view of reading. In Cornoldi, C., and Oakhill, J.V. (Eds.), *Reading comprehension difficulties*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gough, P.B. & Tunmer, W.E. (1986) Decoding, reading and reading disability. *Remedial and Special Education*, 7, 1, 6-10.
- Graesser, A.C., Millis, K.K., & Zwaan, R.A. (1997) Discourse Comprehension. *Annual Review of Psychology*, 48, 163-169.
- Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing Inferences during Narrative Text Comprehension. *Psychological Review*, 101, 371-395
- Graesser, A.C., & Zwaan, R.A. (1995). "Inference Generation and the Construction of Situation Models." In Weaver, C.A., Mannes, S., & Fletcher, C.R. (eds) (1995) "Discourse Comprehension: Essays in Honor of Walter Kintsch." Hillsdale, Lawrence Erlbaum Associates.

- Guthrie, J.T., & Wigfield, A. (2005) Roles of Motivation and Engagement in Reading Comprehension Assessment. In Paris, S.G. & Stahl, S.A., (eds) *Children's Reading Comprehension and Assessment*. Mahwah, New Jersey. Lawrence Erlbaum Associates.
- Gygax, P., Oakhill, J., & Garnham, A. (2003). The representation of characters' emotional responses: Do readers infer specific emotions? *Cognition and Emotion*, 17(3), 413-428.
- Hacker, D.J. (1997). Comprehension monitoring of written discourse across early-to-middle adolescence. *Reading and Writing*, 9, 207-240.
- Hagtvet, B.E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing*, 16, 505-539.
- Harlaar, N., Spinath, F. M., Dale, P. S., & Plomin, R. (2005). Genetic influences on word recognition abilities and disabilities: A study of 7 year old twins. *Journal of Child Psychology and Psychiatry*, 46, 373-384.
- Hatcher, P. (2000). Predictors of Reading Recovery book levels. *Journal of Research in Reading*, 23, 67-77.
- Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. *Reading and Writing*, 2, 127-160.
- Hulme, C., & Snowling, M.J. (2009). *Developmental Disorders of Language Learning and Cognition*. Chichester, John Wiley & Sons.
- Kameenui, E.J., Carnine, D.W., & Freschi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly*, 17, 367-388.

- Keenan, J., Betjemann, R., Wadsworth, DeFries, & Olson (2006). Genetic and Environmental Influences on Reading and Listening Comprehension. *Journal of Research in Reading*, 29, 1, 75-91.
- Kendeou, P., Bohn-Gettler, C., White, M.J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, 31, 3, 259-272.
- Kintsch, W. (1998) *Comprehension*. New York. Cambridge University Press.
- Kintsch, W., & Kintsch, E. (2005) Comprehension. In Paris, S.G. & Stahl, S.A., (eds) *Children's Reading Comprehension and Assessment*. Mahwah, New Jersey. Lawrence Erlbaum Associates.
- Kintsch, W & Rawson, K.A. (2007) Comprehension. In Hulme, C., & Snowling, M.J. (Eds), *The Science of Reading: A Handbook*. Oxford. Blackwell.
- Leach, J.M., Scarborough, H.S., & Rescorla, L. (2003) Late-emerging reading disabilities. *Journal of Educational Psychology*, 95, 211-224.
- Leather, C.V., & Henry, L.A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology*, 58, 88-111.
- Lorch Jnr, R.F., Klusewitz, M.A., & Lorch, E.P. (1995) Distinctions among reading situations. In Lorch Jnr, R.F., & O'Brien, E.J., (eds). *Sources of Coherence in Reading*. Hillsdale NJ, Lawrence Erlbaum Associates.
- Lorch Jnr, R.F., Klusewitz, M.A. & Lorch, E.P. (1993) College students' conditional knowledge about reading. *Journal of Educational Psychology*. 85, 239-252.
- Manis, F., Seidenberg, M., Doi, L., McBride-Chang, C., & Peterson, A., (1996). On the basis of two subtypes of developmental dyslexia. *Cognition*, 58, 157-195.



- Marshall, C.M. & Nation, K. (2003). Individual differences in semantic and structural errors in children's memory for sentences. *Educational and Child Psychology*, 20, 7-18.
- McDougall, S., Hulme, C., Ellis, A. & Monk, A. (1994). Learning to read: The role of short-term memory and phonological skills, *Journal of Experimental Child Psychology* 58: 112-133.
- McGee, A. and Johnson, H. (2003). 'The effect of inference training on skilled and less skilled comprehenders', *Educational Psychology*, **23**, 1, 49-59.
- McKeown, M.G., Beck, I.L., Omanson, R.C. and Perfetti, C. (1983) The effects of long-term vocabulary instruction on reading comprehension: a replication. *Journal of Reading Behavior*, 15. 3-18.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99(3), 440-466.
- McNamara, D.S., Levinstein, I.B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, 36, 222-233.
- McNamara, D.S., & O'Reilly, T. (2009 in press). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In F. Columbus (Ed.), *Progress in Experimental Psychology Research*. Hauppauge, NY: Nova Science Publishers, Inc.
- Mezynski, K. (1983) Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, 53, 253-279.
- Muter, V., Hulme, C., Snowling, M.J. & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40, 663-681.

- Myers, J.L., O'Brien, E.J., Albrecht, J.E., and Mason, R.A. (1994) Maintaining Global Coherence during Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 4, 876-886.
- Nagy, W.E. (1988) *Teaching Vocabulary to Improve Reading Comprehension*. Newark, DE: International Reading Association.
- Nagy, W.E., & Anderson, R.C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19, 304-330.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American educational Research Journal*, 24, 237-270.
- Nash, H., & Snowling, M.J. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language and Communication Disorders*, 41, 335-354.
- Nation, K. (2007) Children's Reading Comprehension Difficulties. In Hulme, C., & Snowling, M.J. (Eds), *The Science of Reading: A Handbook*. Oxford. Blackwell.
- Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, 73, 139-158.
- Nation, K., Clarke, P., Marshall, C.M., & Durand, M. (2004). Hidden language impairments in children: parallels between poor reading comprehension and specific language impairment. *Journal of Speech, Hearing and Language Research*. 47, 199-211.
- Nation, K., Clarke, P., & Snowling, M.J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology*, 72, 549-560.
- Nation, K., Snowling, M.J., & Clarke, P. (2007) Dissecting the relationship between language skills and learning to read. Semantic and phonological contributions to new

- vocabulary learning in children with poor reading comprehension. *Advances in Speech-Language Pathology*, 9, 131-139.
- Nation, K., Cocksey, J., Taylor, J., & Bishop, D.V.M. (2010 in press). A longitudinal investigation of the early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*.
- Nation, K., & Snowling, M.J. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67, 359-370.
- Nation, K., & Snowling, M.J. (1998a). Semantic processing and the development of word recognition skills: evidence from children with reading comprehension difficulties. *Journal of Memory & Language*, 39, 85-101.
- Nation, K., & Snowling, M.J. (1998b). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development*, 69, 996-1011.
- Nation, K., & Snowling, M.J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: evidence from semantic priming. *Cognition*, 70, B1-13.
- Nation, K. & Snowling, M.J. (2000). Factors influencing syntactic awareness in normal readers and poor comprehenders. *Applied Psycholinguistics*, 21, 229-24
- Nation, K., & Snowling, M.J. (2004). Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27, 342-346.
- Nation, K. Snowling, M. J. & Clarke, P.J. (2005). Production of the English past tense by children with and without language comprehension impairments. *Journal of Child Language*, 32, 117-137.

- Neale, M.D. (1997). *Neale Analysis of Reading Ability – Second British Edition (NARA II)*. London, Nfer-Nelson.
- Oakhill, J.V. (1982). Constructive Processes in Skilled and Less Skilled Comprehenders' Memory for Sentences. *British Journal of Psychology*, 73, 13-20.
- Oakhill, J.V. (1983) Instantiation in skilled and less-skilled comprehenders. *Quarterly Journal of Experimental Psychology*, 35a, 441-450.
- Oakhill, J.V. (1984) Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, 54, 31-39.
- Oakhill, J.V. Cain, K., & Yuill, N. (1998) Oakhill, J.V., Cain, K. and Yuill, N.M. (1998) Individual differences in children's comprehension skill: Towards an integrated model. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorder*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Oakhill, J.V., Hartt, J., & Samols, D. (2005), Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, 18, 657-686.
- Oakhill, J.V., Yuill, N., & Parkin, A. (1986). On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading*, 9, 80-91.
- O'Brien, E.J. & Albrecht, J.E. (1992). Comprehension strategies in the development of mental models. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 18, 777-784.
- Ozuru, Y., Dempsey, K., & McNamara, D.S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, 19, 228-242.
- Palincsar, A.S., & Brown, A.L. (1984) Reciprocal teaching of comprehension fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.

- Pany, D., Jenkins, J.R. and Schreck, J. (1982) Vocabulary instruction: effects on word knowledge and reading comprehension. *Learning Disability Quarterly*, 5, 202-215.
- Paris, S.G., & Lindauer, B.K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, 8, 217-227.
- Paris, S.G., & Myers, M. (1981). Comprehension monitoring, memory, and study strategies of good and poor readers. *Journal of Reading Behavior*, 23, 5-22.
- Paris, S.G., & Upton, L.R. (1976). Children's memory for inferential relationships in prose. *Child Development*, 47, 660-668.
- Perfetti, C.A. (1985) "Reading Ability". Oxford: Oxford University Press.
- Perfetti, C.A. (1994) Psycholinguistics and reading ability. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Perfetti, C. A., Landi, N. & Oakhill, J. (2007). The Acquisition of Reading Comprehension Skill. In C. Hulme & M.J. Snowling, (Eds.), *The Science of Reading: A Handbook*. Oxford. Blackwell.
- Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56-115.
- Raaijmakers, J.G.W., Schrijnemakers, J.M.C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41, 416-426.
- Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11, 289-312.
- Raven, J.C. (1998). *Raven's Standard Progressive Matrices (SPM)*. Harcourt Assessment, San Antonio. The Psychological Corporation.

- Ricketts, J., Bishop, D.V.M., & Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *Journal of Research in Reading, 31*, 117-135.
- Ricketts, J., Nation, K., & Bishop, D. V. (2007) Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*, 235-257.
- Rosenshine, S.V. (1980). Skill hierarchies in reading comprehension. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.). *Theoretical Issues in Reading Comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rupley, W.H., Wilson, V.L., & Nichols, W.D. (1998). Exploration of the developmental components contributing to elementary school children's reading comprehension. *Scientific Studies of Reading, 2*, 143-158.
- Seignuric, A.S. & Ehrlich, M.F. (2005) Contribution of Working Memory Capacity to Children's Reading Comprehension: A Longitudinal Investigation. *Reading and Writing, 18*, 7-9, 617-656.
- Seignuric, A.S., Ehrlich, M.F., Oakhill, J.V., & Yuill, N.M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing, 13*, 81-103.
- Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language, 31*, 507-524.
- Snow, C., Tabors, P., & Dickinson. D. (2001). Language development in the preschool years (pp. 1-26). In D. Dickinson, & P. Tabors (Eds.), *Beginning literacy with language: Young children learning at home and school*. Baltimore, MD: Paul H. Brookes Publishing.
- Snowling, M.J. (2000). *Dyslexia*, (2<sup>nd</sup> Edition). Oxford: Blackwell.

- Snowling, M.J., Bishop, D.V.M., & Stothard, S.E. (2000). Is pre-school language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry*, 41, 587-600.
- Snowling, M.J., & Frith, U. (1986). Comprehension in "Hyperlexic" Readers. *Journal of Experimental Child Psychology*, 42, 392-415.
- Snyder, L., Caccamise, D., & Wise, B. (2005) The Assessment of Reading Comprehension: A cautionary tale. *Topics in Language Disorders*, 25, 33-50.
- Soederberg, L.M., & Stine, E.A.L. (1995). Activation of Emotion Information in Text Among Younger and Older Adults. *Journal of Adult Development*, 2(1), 23-36.
- Stanovich, K. (1986). 'Matthew effects in reading: some consequences of individual differences in the acquisition of literacy', *Reading Research Quarterly*, 21, 360-406.
- Stanovich, K.E. (1993). Does reading make you smarter? Literacy and the development of verbal intelligence. In H. Reese (ed.), *Advances in child development and behavior*, 24, San Diego, CA. Academic Press.
- Stothard, S.E., & Hulme, C. (1992). Reading comprehension difficulties in children: the role of language comprehension and working memory skills. *Reading and Writing*, 4, 245-256.
- Stothard, S.E., & Hulme, C. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychology and Psychiatry*, 36, 399-408.
- Stothard, S.E., & Hulme, C. (1996). A comparison of reading comprehension and decoding difficulties in children (pp. 93-112). In Cornoldi, C. and Oakhill, J. (Eds). *Reading comprehension difficulties: Processes and interventions*. New Jersey: Lawrence Erlbaum Associates.

- Stothard, S.E., Snowling, M.J., Bishop, D.V.M., Chipchase, B. & Kaplan, C. (1998).  
Language-impaired pre-schoolers: A follow-up in adolescence. *Journal of Speech, Language and Hearing Research*, 41, 407-418.
- Swanson, H.L., & Berninger, V. (1995). The role of working memory in skilled and less-skilled readers' comprehension. *Intelligence*, 21, 83-108.
- Taylor, R.S., O'Reilly, T., Rowe, M., & McNamara, D.S. (2009) Improving Understanding of Science Texts: iSTART Strategy Training vs a Web Design Control Task.  
Manuscript submitted for publication.
- Torgesen, J.K., Wagner, R. & Rashotte, C. (1999) *Test of Word Reading Efficiency (TOWRE)*. Pearson Assessment, London: The Psychological Corporation.
- Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M.J., & Lorch, E.P. (2005) Assessment of Comprehension Abilities in Young Children. In Paris, S.G. & Stahl, S.A., (eds) *Children's Reading Comprehension and Assessment*. Mahwah, New Jersey. Lawrence Erlbaum Associates.
- Van den Broek, P., Risen, K., & Husebye-Hartmann, E., (1995). The Role of Readers' Standards for Coherence in the Generation of Inferences during Reading. In Lorch Jr, R.F., & O'Brien, E.J., (eds). *Sources of Coherence in Reading*. Hillsdale NJ, Lawrence Erlbaum Associates.
- Van Dijk, T.A., & Kintsch, W. (1983) *Strategies of Discourse Comprehension*. New York. Academic Press.
- Wechsler, D. (1990). *Wechsler Objective Reading Dimensions*. Pearson Assessment, London: The Psychological Corporation.
- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. Harcourt Assessment, San Antonio: The Psychological Corporation.



- Willson, V.L., & Rupley, W.H. (1997). A structural equation model of reading comprehension based on background, phonemic and strategy knowledge. *Scientific Studies of Reading, 1*, 45-63.
- Wise, J.C., Sevcik, R.A., Morris, R.D., Lovett, M.W., & Wolf, M. (2007). The relationship and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language and Hearing Research, 50*, 1093-1109.
- Yuill, N., & Joscelyne, T. (1988). Effects of organizational cues and strategies on good and poor comprehender's story understanding. *Journal of Educational Psychology, 80*, 152-158.
- Yuill, N. & Oakhill, J. (1988). 'Effects of inference awareness training on poor reading comprehension', *Applied Cognitive Psychology, 2*, 1, 33-45.
- Yuill, N., & Oakhill, J.V. (1991) Children's problems in text comprehension. Cambridge: Cambridge University Press.
- Yuill, N., Oakhill, J.V., & Parkin, A. (1989). Working Memory, Comprehension Ability and the Resolution of Text Anomaly. *British Journal of Psychology, 80*, 351-361.
- Young, A., Perrett, D., Calder, A., Sprengelmeyer, R. & Ekman, P. (2002). *Facial Expressions of Emotion – Stimuli and Tests (FEEST)*. Bury St. Edmunds, England: Thames Valley Test Company.
- Zinar, S. (2000) The Relative Contributions of Word Identification Skill and Comprehension-Monitoring Behavior to Reading Comprehension Ability. *Contemporary Educational Psychology, 25*, 363-377.

# **Appendix 1**

## **Experiment 1 Test Materials**

## Appendix 1 – Stories used in Experiment 1

### **1. Happiness**

#### 1.1 Consistent

Emma wanted to be a famous artist when she grew up. At home she drew lots of good pictures. One day, Emma saw an art programme on television. There was a contest to see who could draw the best Easter card. Emma drew a picture of an Easter Egg Hunt. Her mum sent it in to the television company. Two weeks later, there was a knock at the door. Emma had sent in the best picture and had won a big prize. *Emma was very happy indeed.* She could not wait to tell her friends.

#### 1.2 Inconsistent

There was one place left in the school netball team. The team were playing in a big competition on Saturday. Emma loved netball and wanted to play for the team. Her teacher had to decide who else should be in the team. It was between Emma and two other girls. They held hands and waited to find out who had the last place. It was Emma! She had got the last place in the netball team. *Emma was very upset indeed.* She told everyone that she would do her best to help the team win.

#### 1.3 Neutral

Emma was doing a family history project at school. Mum took Emma to see Gran. Emma wanted to ask Gran about her family. Gran told her some stories and gave her some old photographs. She even told Emma about her own Gran, Emma's great, great Gran! Emma wrote everything down in her project, and handed it in. When he gave back the project, Emma's teacher said it was very good. He gave Emma three house points for good work. *Emma did very well indeed.* She had never been given three house points for something before.

## **2. Fear**

### 2.1 Consistent

Holly was on a school trip. Her class were going to the zoo as part of 'wild week'. They saw elephants, monkeys, zebras and tigers. Holly wanted to see the lions most of all. She could see them, but they were all asleep. "Wake up!" shouted Holly. Suddenly, one of the lions jumped up and ran towards her. The lion ran up to the fence where Holly stood and did a huge roar! *Holly had never been so scared before.* She ran away fast. "He didn't want to be woken up!" said her teacher.

### 2.2 Inconsistent

Holly and her family were on holiday. They were staying in a cottage in a big forest. Holly and her brother were playing in the trees. Then, Holly saw something big and brown moving around. "A bear!" thought Holly. She'd heard that wild bears liked to eat children for dinner! The children ran back to the cottage as fast as they could. There was a bear in the forest! *Holly had never been so brave before.* "That's not a bear, silly!" said Dad. "The family in the next cottage have got a big dog!"

### 2.3 Neutral

Holly went for a long walk in the woods after school. She was doing a nature project about trees. The teacher had asked her class to draw leaves of different trees. She had found an oak tree, a beech tree and lots of pine trees. Now it was getting cold and dark. "It must be very late," thought Holly suddenly. "I must get home quickly." She started to run home but could not find the way out. *Holly had never been so late before.* She ran and ran. Holly was lost in the woods.

### **3. Sadness**

#### 3.1 Consistent

Ben put the rest of his things into his bag. He was all packed and ready for his first holiday abroad. They were going to Spain for a week's holiday by the sea. Mum came in with Max the dog, and told Ben to say goodbye to him. "Why? Where is Max going?" asked Ben. "I want him to come with us, he'll like it." "Don't be silly," said Mum. "Dogs are not allowed in aeroplanes. Max is going to stay with Grandma." Ben gave Max a big hug and said goodbye. *Ben was very sad because Max could not go.* He was really going to miss him.

#### 3.2 Inconsistent

It was the morning of the party at last! Ben was going to Alton Towers with his friends. After that they were going out to Pizza Hut for dinner. But that morning Ben felt a bit funny. His tummy, arms and legs were very itchy. He looked down and saw there were red spots all over them. "Mum!" he shouted "I'm all spotty!" Mum ran upstairs to see what was wrong with Ben. "Oh dear, Ben," said Mum. "You have got chickenpox so you will have to stay at home." *Ben was very glad because he could not go.* He'd waited for a long time for this trip.

#### 3.3 Neutral

It was half term. That day, Ben was going to the cinema with his friends. Ben's mum was taking them to see the new Harry Potter film. But, that morning, Mum couldn't stop sneezing! She had a sore throat, a runny nose, and her head hurt. "Ben", said Mum, "I'm not feeling well." "We will have to go to the cinema another day." Ben began ringing his friends to tell them the bad news. "We'll go to the cinema another day," said Ben. "Poor mum will be feeling better then." *Ben was ringing them because he could not go.* He had to tell them the trip was off.

## **4. Disgust**

### 4.1 Consistent

Anna was eating a bowl of salad. She had fresh lettuce, tomato and cucumber from her garden. She put some lettuce in her mouth and started to chew it. Suddenly she spat out the lettuce and screamed. There was something wriggling in the lettuce. Anna had felt it wriggling in her mouth. She looked in the bowl and saw a small green caterpillar. The caterpillar had been hiding in the lettuce. She told mum to wash the lettuce much better in future. *Anna was so disgusted she didn't eat her lunch.*

### 4.2 Inconsistent

Anna and Katie were cleaning out their hamster's cage. Anna picked him up in her hand. She held him carefully so he could not escape. The hamster did not like being held. Katie quickly changed the sawdust at the bottom of the cage. Then she put some fresh straw in his house. The hamster kept trying to get away, then Anna squealed. She put the hamster back in his clean cage very quickly. The hamster had done a poo on her hand. *Anna was so delighted she didn't eat her lunch.*

### 4.3 Neutral

Anna was packing for her holiday. It was half term and the whole family were going to France. Anna had never been there before. She was learning to speak French at school. Her class also learned about French food in cookery week. They talked about French bread and strange cheeses. Then, Anna remembered that in France they like eating snails. Anna wondered if you had to eat the shell as well. And, were they slimy like the ones in her garden? *Anna was so busy packing she didn't eat her lunch.*

## **5. Anger**

### 5.1 Consistent

It was the day after the big test. Jack had worked very hard. He learned all of his spellings and tables. He wanted to be top of the class. In the test Jack had seen Sam copying his answers. He didn't tell his teacher. Sam was a nasty bully and Jack did not want to risk it. The teacher gave them their test results. Jack was second in class with 47 out of 50. "And in first place", said Mrs Green, "is Sam with 48!" Everyone said well done to Sam. *Jack was very angry indeed.* He should have come first, but Sam cheated.

### 5.2 Inconsistent

Mum and dad were getting a new kitchen. They had new cupboards and a new cooker. Then, it was time to paint the new kitchen. Dad opened the paint tin. It was pink paint. Jack hated pink, because he thought it was for girls! "Mum chose it and she is a girl," said Dad laughing. "We are going to paint your room next," said Dad. "We will have lots of pink paint left for it." Jack did not want a pink bedroom. He ran upstairs to his room and slammed the door. *Jack was very happy indeed.* He didn't know that dad was joking.

### 5.3 Neutral

It was Saturday. Jack and Tom had no school, so they went shopping. Jack wanted to buy a train set. He had birthday money to spend. They went to a big toy shop. Tom wanted to buy something but he had no money. He saw some great toys and games. There were some packs of Harry Potter stickers on the shelf. Tom picked up two packs and put them in his pocket. "What are you doing Tom?" said Jack. "Put those back or we will be in big trouble." *Jack was very clever indeed.* He saw that Tom was going to steal the stickers.

## Experiment 1 Comprehension Questions

For each story, the two questions requiring literal information are listed first, followed by the question requiring a coherence inference, and the question requiring an elaborative inference.

### 1. Happiness

#### 1.1 Consistent

What did Emma want to be when she grew up?

What did Emma draw?

How did Emma find out that she had won?

Who knocked on the door?

#### 1.2 Inconsistent

How many places were left on the netball team?

Who had to decide who got the last place?

What game was the team playing in the competition?

Why did Emma get chosen to play for the team?

#### 1.3 Neutral

What did Gran give to Emma?

How many house points did Emma get for the project?

Why did Emma go and see Gran?

Was it normal to get three housepoints for one piece of work?



## 2. Fear

### 2.1 Consistent

Who were asleep?

What did the lion do after Holly shouted?

Why did the lion roar at Holly?

What do you think “Wild Week” was about?

### 2.2 Inconsistent

Where were Holly and her family staying?

What did Holly think that wild bears liked to do?

Why did Holly run as fast as she could?

How did Dad know it was not a bear that Holly had seen?

### 2.3 Neutral

What was Holly doing a project about?

What trees had Holly found so far?

How did Holly know it was late?

What did Holly take into the woods with her?

### 3. Sadness

#### 3.1 Consistent

Where was Ben going on holiday?

What was the dog's name?

Why was Ben going to miss Max?

What had Ben packed in his bag?

#### 3.2 Inconsistent

Where was Ben going with his friends?

What did Ben have wrong with him?

Why did Ben shout his mum?

Why did having chickenpox mean that Ben couldn't go out?

#### 3.3 Neutral

Who was going to the cinema?

What was Ben going to see at the cinema?

Why didn't Ben have school?

What was mum going to do instead of taking Ben and his friends to the cinema?

## 4. Disgust

### 4.1 Consistent

What was Anna eating?

Where were the lettuce, tomato and cucumber from?

What was wriggling in Anna's mouth?

When did the caterpillar climb into the lettuce?

### 4.2 Inconsistent

What were Anna and Katie doing?

Who picked up the hamster?

Why did Anna put the hamster back in the cage so fast?

Why were Anna and Katie cleaning out the hamster's cage?

### 4.3 Neutral

Where were Anna and her family going?

What foods had they talked about at school?

Why wasn't Anna at school?

What was Anna packing?

## 5. Anger

### 5.1 Consistent

Who came top in the test?

What mark did Jack get in the test?

Why did everyone say well done to Sam?

Did Sam deserve to come top in the test?

### 5.2 Inconsistent

What room were mum and dad going to paint next?

Why did Jack hate pink so much?

Why didn't Jack want them to paint his room?

What were they going to paint in the kitchen?

### 5.3 Neutral

What did Jack want to buy?

How much money did Tom have?

Why did Tom put the stickers in his pocket?

Who did Jack think they would be in big trouble with?

## **Appendix 2**

### **Experiment 2 Test Materials**

## Experiment 2 Stories

### 1. Ball size inconsistency

#### 1.1 Consistent

It was Saturday and it was bright and sunny! Jack knew that meant only one thing.

With a big grin on his face, he got dressed for the beach. He put on his shorts, a t-shirt and his cool shades. Then he picked up his surfboard. He put it in the boot of the car. Dad kicked the football over to Jack. Jack picked up the ball and put it in his bag. He put his bag in the boot and got in the car. “Surf’s up!” said Dad, and off they went.

#### 1.2 Inconsistent

Cup final day at last! Jack had been awake since 5 o’clock. He was goalkeeper in the York under tens football team. They were playing Leeds under tens at Elland Road, home of Leeds United. Some of the Leeds United first team were going to be there. Jack put on his tracksuit and packed his kit. Outside Dad kicked the football to Jack. Jack picked up the ball and put it in his pocket. He put his bag in the boot and got in the car. It was time to go at last.

#### 1.3 Neutral

Jack was a superb tennis player. Since he was three, he had played every day. Now all his hard work was about to pay off. He was one point away from the biggest win of his life. Just one more point to win the trophy. And one more point for a place in the top national competition. He took some deep breaths. Jack picked up the ball and threw it into the air. He hit it as hard as he could over the net. It was a clean winner – Jack was the champion!

## 2. Sofa size inconsistency

### 2.1 Consistent

It was Harry's birthday. He had one hundred pounds of birthday money to spend. He wanted the new England World Cup football shirt. He also wanted a computer game and a DVD. Harry's birthday was in the Easter holidays. As a special treat, Dad took the day off work. Dad took Harry into York on the bus. In a sports shop Harry tried on the new England shirt. It was a perfect fit. Dad picked up the shirt and put it in the shopping basket. They went to the checkout and paid for it. Next on Harry's shopping list was the computer game.

### 2.2 Inconsistent

Harry was bored and hungry. He was shopping with mum and dad. They were looking for a new sofa. They had been to all of the sofa shops in York. Mum and Dad were standing by a big red sofa. They had been talking about it for ages. "Do you like this sofa Harry?" asked Mum. "It's alright," mumbled Harry. "We will buy it," said Dad "it is the best one we have seen". Dad picked up the sofa and put it in the shopping basket. They went to the checkout and paid for it. Now they could go and get some lunch.

### 3. Bowl size inconsistency

#### 3.1 Consistent

Emma and her friend went for a walk in the woods. After two hours, it was time to turn back. They wanted to get home for lunch. They tried to go back the same way but took a wrong turn. They were lost. In the end Emma's dad found them. It was nearly dinner time when they got home. They were very tired and hungry. Mum gave them some crisps while they waited for dinner. Emma got some from the bowl and started to eat. They were the nicest crisps she had ever tasted. She was very glad to be home.

#### 3.2 Inconsistent

Emma and mum were making a chocolate cake. Emma mixed up some butter and sugar. Mum added an egg. Then they mixed in some flour and baking powder. Next, they melted some chocolate in a bowl. Mum poured the chocolate into the cake mix. There was some left on the sides of the bowl. "You can eat the rest of the chocolate, Emma" said mum. Mum knew how much Emma loved chocolate. Emma got into the bowl and started to eat. It was the nicest chocolate she had ever tasted. She did not stop eating until the bowl was clean.

#### 3.3 Neutral

Emma, mum and dad got to the hotel at last. It had taken ten long hours to get there! They had travelled by car, aeroplane, and coach. Now they could relax and enjoy the holiday. Mum unpacked all of their things. After the long hot journey they wanted to go for a swim. The hotel had a big pool in the gardens. They put on their sun cream and swimming costumes. Then they went outside. Emma jumped into the pool and started to swim. It was the nicest pool she had ever been in. The long journey had been worth it!



#### 4. Race time inconsistency

##### 4.1 Consistent

Katie was the best runner in her school. She was a member of the athletics club. She dreamed of running in the Olympics in 2012. Her best event was the eight hundred metres. On sports day, she ran in lots of races. She won the one hundred metres, and the hurdles. Her next race was the eight hundred metres. She wanted to break the school record. She waited to start. In a few minutes Katie would be at the finish line. Then she could have a rest and cool down.

##### 4.2 Inconsistent

Katie was on a ten mile fun run. She was raising money for Children in Need. Lots of her family and friends were sponsoring her. She had been running for an hour. It was a very hot day and Katie was tired. Her legs had started to hurt. She stopped to get a drink of water. She poured some cold water over her head. There were four miles to go. In a few seconds Katie would be at the finish line. Then she could have a rest and cool down.

## Experiment 2 Comprehension Questions

For each story, the two questions requiring literal information are listed first, followed by the question requiring a coherence inference, and the question requiring an elaborative inference.

### 1. Ball size inconsistency

#### 1.1 Consistent

What day of the week was it?

What was the weather like?

What were they going to do at the beach?

Where did Jack live?

#### 1.2 Inconsistent

What position did Jack play in football?

Which team did Jack play for?

Where were Jack and his dad going?

Why did Jack wake up so early?

#### 1.3 Neutral

What game was Jack superb at?

What happened if Jack won the next point?

What was Jack about to do?

How did Jack beat his opponent to win the point?

## 2. Sofa size inconsistency

### 2.1 Consistent

How much birthday money did Harry have?

How did they get to York?

Who went shopping with Harry?

In what month did Harry celebrate his birthday?

### 2.2 Inconsistent

Where had they been shopping?

What did they want to buy?

What had they been talking about for ages?

Where were they going to put the sofa?

### 3. Bowl size inconsistency

#### 3.1 Consistent

Who went for a walk?

Who found them?

Why did they turn back after 2 hours?

Why were the crisps the nicest Emma had ever tasted?

#### 3.2 Inconsistent

What sort of cake were they making?

What did they do with the chocolate?

Tell me 3 things they put in the cake.

How did they melt the chocolate?

#### 3.3 Neutral

Where was the swimming pool?

How had they travelled there?

Why did they put on sun cream?

What was the weather like?

#### 4. Race Time Inconsistency

##### 4.1 Consistent

What was Katie's best event?

What was Katie's dream?

When did she win the one hundred metres and hurdles?

Why did Katie want to break the school record?

##### 4.2 Inconsistent

How long was the fun run?

Why did Katie stop?

How was she raising money for Children In Need?

What would the sponsor money be used for?

## **Appendix 3**

### **Experiment 3 Test Materials**

## **Stories – condition 1: Early-acquired stories**

### **1. Set 1**

#### 1.1 Consistent

Jack had a very sore throat and a high temperature.  
He had been off school for three days, and he still wasn't feeling better.  
"I think we need to get the doctor to have a look at you" said Mum.  
Mum phoned the surgery.  
Later that afternoon Mum and Dad took Jack to see the doctor.  
He asked Jack to open his mouth and say "aaah".  
He took a look at Jack's throat.  
"That looks sore" said the doctor.  
"I will give you an injection in your arm now and some tablets to take at home."  
He did not cry when the doctor stuck the needle in.  
The doctor said he would see Jack again next week.  
He wrote out a prescription, and gave it to mum.  
"Take one tablet three times a day" he told Jack.  
Mum made an appointment for a week's time.  
Jack hoped that he would feel much better by then.

#### 1.2 Inconsistent

Jack and dad took their cat, Oscar, to see the vet.  
Once a year, Oscar needed some vaccinations.  
These were special injections that stopped him from getting nasty diseases.  
He was very cross when they tried to put him in the cat basket.  
He put his claws out and wriggled furiously.  
Once he was safely in the basket, they put him in the car.  
All the way to the vets, he meowed at the top of his voice.  
When they got there, they took Oscar into the vet's office.  
The vet examined him, and gave him an injection.  
He did not bark when the vet stuck the needle in.  
The vet said he would see Oscar again next year.  
Jack and dad took Oscar home.  
He did not make a sound on the way back.  
"I think he's glad that the visit to the vet is over," said Jack.

## **2. Set 2**

### **2.1 Consistent**

Emily was having a riding lesson.  
She was learning about jumping.  
The riding teacher got them to make the ponies canter around the paddock.  
Then she set up two jumps for them.  
The pony that Emily was riding refused one of the jumps, and Emily fell off.  
She landed awkwardly on the jump, and screamed with pain.  
Her leg was very sore indeed.  
After a few minutes, she tried to get up.  
It was so painful she could not walk on it.  
One of Emily's friends rushed off to find her mum.  
In the end she was taken to hospital in an ambulance.  
She had broken her leg, and they put it in plaster.  
She needed to keep the plaster on for six weeks.  
Now she understood why it hurt so much!

### **2.2 Inconsistent**

Tom was in the park with his friends.  
There were swings, a slide and a climbing frame.  
Tom wanted to show his friends how good he was at climbing.  
He got to the top of the climbing frame and waved to them.  
They waved back and smiled at him.  
Then Tom slipped, and fell to the ground.  
"Owww!!!" he cried and he curled up in a ball.  
He landed on his arm, and it really hurt.  
It was so painful he could not walk on it.  
One of Tom's friends rushed off to find his mum.  
Mum came running over to Tom.  
She took a look at Tom's arm.  
"I think we need to take you to the hospital" said Mum.  
"We'll get the doctor to check that you haven't broken anything."



### **3. Set 3**

#### **3.1 Consistent**

Yesterday was Rachel's birthday and she had three friends around for a sleepover. They had watched a couple of films, and they played on Rachel's playstation. For her birthday she had got a game called Singstar. They had to sing along to pop records, and the playstation worked out who sang the best. The girls had a lot of fun. They went to bed very late. The next morning they had a lie in. They got up and put on their dressing gowns. They went downstairs to get their breakfast. There were some great smells coming from the kitchen. Mum had made bacon and eggs for everyone.

#### **3.2 Inconsistent**

The ring of her alarm clock told Rachel it was time to get up. She pressed the snooze button and rolled over. Rachel hated getting up early, and she did not want to get out of bed. Her brother Ben had other ideas. He barged into her bedroom and splashed her with cold water. "Ben, GO AWAY!" moaned Rachel. "Time to get up, lazy bones! We're going to Flamingo Land, remember?" said Ben. Rachel rolled out of bed and put on her dressing gown. She went downstairs to get her dinner. There were some great smells coming from the kitchen. Mum had made bacon sandwiches for everyone.

## **Condition 2 Late-Acquired**

### **4. Set 4**

#### 4.1 Consistent

Sam went to a car boot sale with his dad.  
They were trying to sell some of their old stuff and make a bit of money.  
After they had sold most of their things, Sam went for a walk.  
He wanted to buy a birthday present for mum with the money he made.  
He found somebody selling fridge magnets.  
Mum liked cats, so he bought a set with kittens on them.  
He took them back to show dad.  
He picked up some coins with the magnets.  
Then he picked up a spoon.  
Sam explained they would stick to the door of the fridge.  
And mum would see them whenever she was in the kitchen.  
Dad thought mum would be very happy with them.

#### 4.2 Inconsistent

Jack and Ben were working together in a science lesson.  
They were learning about magnetic forces.  
The teacher had given them a strong magnet.  
They had to try and pick things up with it.  
Jack picked up a nail, and a pair of scissors.  
Ben picked up some paper clips and his pencil tin, but he couldn't pick up his pencil.  
Jack got his packed lunch out of his school bag.  
He picked up his apple with the magnet.  
Then he picked up his spoon.  
Ben looked around for something else made of metal that the magnet would pick up.  
There was a stapler on the teacher's desk.  
The magnet stuck to the stapler, but it wasn't strong enough to pick it up.

## **5. Set 5**

### **5.1 Consistent**

Molly was helping dad to clear the vegetable patch.  
They were getting rid of all of the weeds and sowing some new seeds.  
They were growing carrots, potatoes, leeks and cabbages.  
Dad had also made a frame so they could grow some runner beans.  
While Molly was digging she found something very strange hanging from one of the weeds.  
It looked like a bag with a little alien in it!  
She called dad over to have a look.  
Dad explained that it was a cocoon, and that a caterpillar had made it.  
He told her all about how caterpillars turn into butterflies.  
Molly was amazed that they could change like that.

### **5.2 Inconsistent**

Molly and Emma were outside looking for butterflies and caterpillars.  
In Science, they were learning about insects.  
In the school garden they had buddleia plants.  
Molly and Emma knew that butterflies loved the flowers on them.  
They were sure to find a few different coloured butterflies there.  
They also thought they might find caterpillars eating the cabbages in the school vegetable garden!  
Their teacher had told them about the life cycle of butterflies.  
They went through a process called metamorphosis.  
He showed them that butterflies and caterpillars were the same creatures.  
He told them all about how butterflies turn into caterpillars.  
The children were amazed that they could change like that.

## **6. Set 6**

### **6.1 Consistent**

Emily had a pet tortoise called Fred.  
Her friends thought he was a strange pet, but Emily loved him.  
She couldn't have a dog or a cat because she was allergic to animal fur.  
One day, Emily noticed that Fred wasn't moving around in his run.  
The next day, he did not come out again.  
"Mum!" she cried in alarm "I think Fred has died!"  
Mum had a look at Fred.  
"Emily, I think he's only sleeping – when it gets cold, some tortoises hibernate".  
Emily breathed a sigh of relief.  
She knew that some creatures need to hibernate to survive the winter months.  
When the weather gets warmer they wake up.  
She was going to miss him but she was glad he was alright!

### **6.2 Inconsistent**

It was a nice clear evening.  
Megan was watching a fat hedgehog in her garden.  
It was busily looking around and eating bits of food.  
Mum explained to Megan that the hedgehog was eating lots of extra food.  
It needed to put on lots of weight before it could hibernate.  
It needed to get a big store of energy as it would be asleep for a long time.  
Megan remembered that she had learned about this at school.  
In science they learned that bears, hedgehogs and dormice did this every year.  
She knew that some creatures need to hibernate to survive the summer months.  
When the weather gets warmer they wake up.  
She thought it must be weird to be asleep for so long!

### Experiment 3 Comprehension Questions

For each story, the two questions requiring literal information are listed first, followed by the question requiring a coherence inference, and the question requiring an elaborative inference.

Set 1:

Consistent

What did the doctor give to mum?

\*How many tablets did Jack have to take every day?

Why did the doctor ask Jack to open his mouth?

How did you know Jack was brave?

Inconsistent

What did the vet give Oscar?

When did the vet say he would see Oscar again?

Why did Oscar get cross?

What did Oscar do on the way home?

Set 2:

Consistent:

What part of her body did she hurt?

\*How did she travel to hospital?

Why did Emily fall off?

Why was her leg put in plaster?

Inconsistent

What part of his body did Tom land on?

What things were there to play on in the park?

What did Tom fall from?

Why did Mum take Tom to the hospital?

Set 3:

Consistent

What did Rachel get for her birthday?

How many friends did Rachel have over on her birthday?

What did they have for breakfast?

Why did they have a lie in?

Inconsistent

What did Ben splash Rachel with?

Where were they going?

What did the kitchen smell of?

What was Rachel going to have to eat?

Set 4:

Consistent

What did Sam want to buy with the money he made?

What was on the set of fridge magnets Sam bought for mum?

How did Sam make some money?

Why did Sam think that the magnet would stick to the fridge?

Inconsistent

What lesson were they having?

Tell me 3 things the boys picked up with the magnet.

\*Why couldn't the magnet pick up the stapler?

Why did Ben look for something made of metal?

Set 5:

Consistent

What part of plants do butterflies love?

Where did Molly and Emma look for caterpillars?

What were Molly and Emma looking for?

What is metamorphosis?

Inconsistent

What do caterpillars turn into?

Tell me 3 things they were going to grow?

Why were they getting rid of weeds?

What was inside the cocoon?

Set 6:

Consistent

What was the tortoise's name?

When do some tortoises hibernate?

Why was Emily going to miss Fred?

Why was a tortoise a good pet for Emily?

Inconsistent

What was Megan watching?

What other animals hibernate?

What was the hedgehog looking for?

Why was the hedgehog fat?

\* for Experiment 4 these questions were amended as the content of the stories had changed slightly when post-target sentences were added. For Experiment 4 the questions asked were

“What did Jack have to take at home?”

“Where did mum think she needed to take Jo?”

“Why couldn't the magnet pick up the keys?”

**Appendix 4**  
**Experiment 4 Test Materials**



## **Stories – condition 1: Early-acquired stories**

### **1. Set 1**

#### 1.1 Consistent

Jack had a very sore throat and a high temperature.  
He had been off school for three days, and he still wasn't feeling better.  
"I think we need to get the doctor to have a look at you" said Mum.  
Mum phoned the surgery.  
Later that afternoon Mum and Dad took Jack to see the doctor.  
He asked Jack to open his mouth and say "aaah".  
He took a look at Jack's throat.  
"That looks sore" said the doctor.  
"I will give you an injection in your arm now and some tablets to take at home."  
He did not cry when the doctor stuck the needle in.  
They all thought Jack was very brave.  
The doctor said he wanted to see Jack again next week.  
He told them that Jack might need another injection then.  
Mum and dad took Jack home.  
He did not make a sound on the way back.  
"I think he's glad that the visit to the doctor is over" said Dad.

#### 1.2 Inconsistent

Jack and dad took their cat, Oscar, to see the vet.  
Once a year, Oscar needed some vaccinations.  
These were special injections that stopped him from getting nasty diseases.  
He was very cross when they tried to put him in the cat basket.  
He put his claws out and wriggled furiously.  
Once he was safely in the basket, they put him in the car.  
All the way to the vets, he meowed at the top of his voice.  
When they got there, they took Oscar into the vet's office.  
The vet examined him, and gave him an injection.  
He did not bark when the vet stuck the needle in.  
Jack thought Oscar was very brave.  
The vet said he wanted to see Oscar again next year.  
Jack and dad took Oscar home.  
He did not make a sound on the way back.  
"I think he's glad that the visit to the vet is over," said Jack.

## **2. Set 2**

### **2.1 Consistent**

Jo was having a riding lesson.  
She was learning about jumping.  
The riding teacher got them to make the ponies trot around the field.  
Then she set up two jumps for them.  
The pony that Jo was riding refused one of the jumps, and Emily fell off.  
She landed awkwardly on the jump, and screamed with pain.  
Her leg was very sore indeed.  
After a few minutes, she tried to get up.  
It was so painful she could not walk on it.  
One of Jo's friends rushed off to find her mum.  
Mum came running over to Jo.  
She took a look at Jo's leg.  
"I think we need to take you to the hospital" said Mum.  
Jo had broken her leg, and they put it in plaster.  
Now she understood why it had hurt so much!

### **2.2 Inconsistent**

Tom was in the park with his friends.  
There were swings, a slide and a climbing frame.  
Tom wanted to show his friends how good he was at climbing.  
He got to the top of the climbing frame and waved to them.  
They waved back and smiled at him.  
Then Tom slipped, and fell to the ground.  
"Owww!!!" he cried and he curled up in a ball.  
He landed on his arm, and it really hurt.  
It was so painful he could not walk on it.  
One of Tom's friends rushed off to find his mum.  
Mum came running over to Tom.  
She took a look at Tom's arm.  
"I think we need to take you to the hospital" said Mum.  
"We'll get the doctor to check that you haven't broken anything."  
Better safe than sorry, Tom!

### **3. Set 3**

#### 3.1 Consistent

Yesterday was Rachel's birthday and she had three friends around for a sleepover. They had watched a couple of films, and they played on Rachel's playstation.

For her birthday she had got a game called Singstar.

They had to sing along to pop records, and the playstation worked out who sang the best.

The girls had a lot of fun.

They went to bed very late.

The next morning they had a lie in.

They got up and put on their dressing gowns.

They went downstairs to get their breakfast.

There were some great smells coming from the kitchen.

Mum had made bacon sandwiches for everyone.

They sat around the table.

Rachel asked for some brown sauce.

Then she ate her sandwich hungrily.

"Thanks Mum! That was yummy!" said Rachel.

#### 3.2 Inconsistent

The ring of her alarm clock told Rachel it was time to get up.

She pressed the snooze button and rolled over.

Rachel hated getting up early, and she did not want to get out of bed.

Her brother Ben had other ideas.

He barged into her bedroom and splashed her with cold water.

"Ben, GO AWAY!" moaned Rachel.

"Time to get up, lazy bones!"

"We're going to Flamingo Land, remember?" said Ben.

Rachel rolled out of bed and put on her dressing gown.

She went downstairs to get her dinner.

There were some great smells coming from the kitchen.

Mum had made bacon sandwiches for everyone.

They sat around the table.

Rachel asked for some tomato sauce.

Then she ate her sandwich hungrily.

"Thanks Mum! That was yummy!" said Rachel.

## **Condition 2 Late-Acquired**

### **4. Set 4**

#### 4.1 Consistent

Sam went to a car boot sale with his dad.  
They were trying to sell some of their old stuff and make a bit of money.  
After they had sold most of their things, Sam went for a walk.  
He wanted to buy a birthday present for mum with the money he had made.  
He found somebody selling fridge magnets.  
Mum liked cats, so he bought a set with pictures of kittens on them.  
He took them back to show dad.  
He picked up some coins with the magnets.  
Then he picked up a spoon.  
Sam looked around for something else made of metal that the magnets would pick up.  
There was a bunch of keys on dad's table.  
The magnets stuck to the keys but they were not strong enough to pick them up.

#### 4.2 Inconsistent

Sam and Ben were working together in a science lesson.  
They were learning about magnetic forces.  
The teacher had given them a strong magnet.  
They had to try and pick things up with it.  
Sam picked up a nail, and a pair of scissors.  
Ben picked up some paper clips and his pencil tin, but he couldn't pick up his pencil.  
Sam got his packed lunch out of his school bag.  
He picked up his apple with the magnet.  
Then he picked up his spoon.  
Ben looked around for something else made of metal that the magnet would pick up.  
There was a bunch of keys on the teacher's desk.  
The magnet stuck to the keys, but it wasn't strong enough to pick it up.

## **5. Set 5**

### **5.1 Consistent**

Molly was helping dad to clear the vegetable patch.  
They were getting rid of all of the weeds and sowing some new seeds.  
They were growing carrots, potatoes, leeks and cabbages.  
Dad had also made a frame so they could grow some runner beans.  
While Molly was digging she found something very strange hanging from one of the weeds.  
It looked like a bag with a little alien in it!  
She called dad over to have a look.  
Dad explained that it was a cocoon, and that a caterpillar had made it.  
He told her all about how caterpillars turn into butterflies.  
Molly was amazed that they could change like that.  
She found a big hairy caterpillar in the garden.  
She wondered how many cabbage leaves it had eaten to get so big.

### **5.2 Inconsistent**

Molly and Emma were outside looking for butterflies and caterpillars.  
In Science, they were learning about insects.  
In the school garden they had lots of plants.  
Molly and Emma knew that butterflies loved the flowers on them.  
They were sure to find a few different coloured butterflies there.  
They also thought they might find caterpillars eating the cabbages in the school vegetable garden!  
Their teacher had told them about the life cycle of butterflies.  
They went through a process called metamorphosis.  
He showed them that butterflies and caterpillars were the same creatures.  
He told them all about how butterflies turn into caterpillars.  
The children were amazed that they could change like that.  
They found a big hairy caterpillar in the garden.  
They wondered how many cabbage leaves it had eaten to get so big.

## **6. Set 6**

### **6.1 Consistent**

Emily had a pet tortoise called Fred.

Her friends thought he was a strange pet, but Emily loved him.

She couldn't have a dog or a cat because she was allergic to animal fur.

One day, Emily noticed that Fred wasn't moving around in his run.

The next day, he did not come out again.

"Mum!" she cried in alarm "I think Fred has died!"

Mum had a look at Fred.

"Emily, I think he's only sleeping – when it gets cold, some tortoises hibernate".

Emily breathed a sigh of relief.

She knew that some creatures need to hibernate to survive the winter.

When the weather gets warmer they wake up.

She thought it must be odd to stay asleep for so long.

"Animals must dream a lot when they hibernate", thought Emily. "Or they would get bored of sleeping."

### **6.2 Inconsistent**

It was a nice clear evening.

Megan was watching a fat hedgehog in her garden.

It was busy looking around and eating bits of food.

Mum explained to Megan that the hedgehog was eating lots of extra food.

It needed to put on lots of weight before it could hibernate.

It needed to get a big store of energy as it would be asleep for a very long time.

Megan remembered that she had learned about this at school.

In science they learned that bears, hedgehogs and dormice did this every year.

She knew that some creatures need to hibernate to survive the summer.

When the weather gets warmer they wake up.

She thought it must be weird to be asleep for so long!

"Animals must dream a lot when they hibernate," said Megan. "Or they would get bored of sleeping."

**Appendix 5**  
**Experiment 6 Test Materials**

## **Experiment 6 - Item Development: Generation of responses to test items**

### **Instructions to participants**

**Thank you for agreeing to take part in this study.**

### **Instructions:**

On the sheets which follow, you will be presented with a series of two or three sentence scenarios. After each scenario a question is posed. Your task is to provide answers to each question that are very likely to be correct, quite likely to be correct, quite unlikely to be correct and very unlikely to be correct. For example given the scenario:

“Tom kicked the ball to the striker. He took it around the defender and decided to try a shot at the goal.”

And the question: “What game was Tom playing?”

It is **very likely** that Tom was playing football because he kicked the ball, there was a striker and a defender, and the striker shot at a goal. It is **quite likely** that he was playing rugby, although the fact that there was a striker and a defender on the pitch suggest otherwise. It is **quite unlikely** that he was playing hockey, as you don’t kick the ball in hockey. And let’s face it, it is **very unlikely** that he was playing Scrabble.

So, in short, I’d like you to suggest answers to the questions that vary in plausibility. Please note that there is no right or wrong answer to these items, I am looking for the most commonly occurring suggestions for each one.

Please do not spend too much time thinking about each item – and if you are really stuck, move on to the next one and leave the answer blank.

When you have finished, please return the consent form and question form to me via the box in the departmental office by Thursday 1 March 2007. I will then issue your subject hour form and put that in your pigeonhole when I collect the forms on Thursday.

Please note that this study is for 1 subject hour, so please spend no more than 1 hour on this task.

Thanks for your help!  
Leesa Clarke  
l.clarke@psych.york.ac.uk



1. Sue asked for an orange juice and lemonade. Jack ordered a diet coke with no ice.

Where were they?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

2. At last it was time for dinner. Tigger was hungry and ate all of the food in his bowl. He curled up in his basket and licked his paws.

What was Tigger

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

3. Mum asked the woman if she could see what was wrong with Nathan's eye. It was very red, itchy and sore. She thought he might need some cream to make it better.

Where were they?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

4. The cat did not like being put into the box and made lots of noise. Mum shut the box, and carried it to the car.

Where were they going?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

5. Anna said she wanted it shorter for the summer. The lady picked up her scissors and began to cut.

Where was Anna?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

6. Mum asked for the cod and chips. Dad wanted the haddock with mushy peas.

Where were they?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

7. Jack was at work. He opened the book at the right page and picked up the trumpet.

What was Jack's job?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

8. They stood in a line, waiting for the gun to fire. Then it was time to run!

What were they doing?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

9. The music was so loud they couldn't hear themselves speak. They pushed their way through the crowd and tried to get to the front.

Where were they?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

10. Callum had never seen so many books. There were people working and no-one made a sound.

Where was Callum?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

11. Mum pressed the “on” button and the music started playing. She turned up the volume and started to sing.

What did Mum switch on?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

12. Jack felt terrible, and went to lie down under the duvet. His tummy hurt and all he wanted to do was sleep.

Where was Jack?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

13. The thunder was very loud and the lightning flashed all around. Sarah hid under her blanket and waited for the storm to pass.

How did Sarah feel?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

14. John rode as fast as he could, but he was not going to make it up the hill. In the end he stopped for a rest, then he pushed it the rest of the way to the top.

What was John riding?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

15. The sun was very hot indeed. Holly had forgotten to bring her sun cream, and she was worried that she was going to burn. She covered herself up in her towel.

Where was Holly?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

16. Bob undid the screws and opened the back of the television. He could see straight away why they could not switch it on. He got his tools and started to fix the wires.

What did Bob do for a living?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

17. It was mum's 40<sup>th</sup> birthday and Dad had planned a surprise for her. He booked a big room at a hotel, and had invited all her friends.

What sort of surprise had dad planned?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

18. John was miserable. His throat hurt, his nose was running and he had a bad cough. Mum made him some honey and lemon to drink.

What was wrong with John?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

19. Charlotte could hear something rustling in the bushes. Then she saw a stripey ginger tail poking out from underneath.

What was hiding in the bush?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

20. Ali had been running for a long time. He was in front, but he was getting tired. He still had three miles to go before he could have a rest.

What was Ali taking part in?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

21. It was a hot summer's day and Hannah was weeding in her garden. She could hear something buzzing and it was getting louder. She was scared and looked around in a panic.

What did Hannah think the buzzing was?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

22. The doorbell rang, and a man in a red jacket waited at the door. The parcel was too big to fit through the letterbox, and he needed someone to sign for it.

Who was at the door?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

23. They all stood around the table and sang while Connor blew out the candles. Then it was time to eat the cake his mum had made, and to start playing musical chairs.

Where were they all?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

24. The warm water splashed down on Holly's head. She got some shampoo and rubbed it into her hair.

Where was Holly?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

25. John could see smoke coming from the house. He got his mobile phone and dialed '999'.

Who was John phoning?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

26. John fell over and tore a big hole in the knee of his new jeans. Mum got a needle and thread.

What was mum going to do?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

27. It was time for their summer holiday at last. The whole family was going to Disneyland in Florida, and they were on their way at last.

How were they travelling to Florida?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

28. It was late on Friday night, and Emma was having a party. Suddenly all the lights went out and the music stopped. Everyone looked around at each other in surprise.

What had happened?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

29. The nurse came in pushing a trolley. She told Mum that it was time for her to take some tablets.

Where were they?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

30. Abi was tired and hungry, and the sun was getting hot. They had been walking since breakfast and she had had enough. Dad said it was another 2 miles to the nearest village.

Where were they?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

31. John was really pleased with himself – everyone kept telling him how good he looked and his trousers were getting very baggy.

Why was John pleased?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

32. They had been waiting for over an hour to see the elephants. They came quite close to the car and Freddie couldn't believe how big they were!

Where was Freddie?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

33. Adam put his foot down and turned the wheel. Everybody laughed as he crashed into Mum and Dad. Dad put his foot down and went after Adam to get his own back.

Where was Adam

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

34. It was a hot summer's day and Ella was helping dad to sell some of her old toys. A boy came up to the table, and asked her if there was anything he could buy for a pound.

Where was Ella?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

35. Bob was really annoyed. All of the leaves of his lettuces were gone and the cabbages were half eaten.

What had eaten them?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

36. Tom clicked the mouse and waited for the page to load. He hoped that it would help him with his history homework.

What was Tom using for his homework?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

37. It was very cold and there was frost on the ground. Jack wondered if there would be school tomorrow, because the weather forecast was for heavy showers.

What was expected to fall overnight?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely



38. John sealed the envelope and put a stamp on the front. He hoped that his friend would be glad to hear all his news.

What was John sending?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

39. The referee blew his whistle. The ball was in the back of the net, and the home team was one up.

What were they playing?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

40. The car was making a terrible noise, and Ryan decided that it was not safe to carry on driving. He decided to pull over and ring for help.

Who was Ryan ringing?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

41. After Emily's dad had finished his speech, Ben and Emily cut the cake while everyone took pictures. It had been a wonderful day!

What happened that day?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

42. Clare held her nose and jumped in. The water was freezing and she let out a yell. Mum sat on the side with her feet in the cold water.

Where were they?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

43. It was their first night in the tent. The rain was pouring down and Ben's sleeping bag was getting wet. He wanted to go in the house but thought they would laugh at him.

Where was the tent?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

44. Oliver had been feeling really sick, and he had a very high temperature. Mum decided to ring and see if someone could come and see him.

Who did mum ring?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

45. The man was carrying a bag of jewellery and lots of cash when the police stopped him. The watches and the rings had been stolen and did not belong to him.

Who was the man?

Very likely  
Quite likely  
Quite unlikely  
Very unlikely

Thank you for your help – please return the completed sheets to the box in the departmental office.

## **Experiment 6 - Item Generation: Answer plausibility rating study**

### **Instruction to Participants**

**Thank you for agreeing to take part in this study.**

### **Instructions:**

On the sheets which follow, you will be presented with a series of two or three sentence scenarios. After each scenario a question is posed. Your task is to rank the plausibility of the answers provided from 1 to 4. The example below should give you the idea of what you need to do.

#### **The scenario:**

“Tom kicked the ball to the striker. He took it around the defender and decided to try a shot at the goal.”

The question: “What game was Tom playing?”

Possible answers

Rugby	<b>2</b>	More plausible than hockey...
Darts	<b>4</b>	i.e. the least plausible
Football	<b>1</b>	i.e. the most plausible
Hockey	<b>3</b>	More plausible than darts..

Please note that there is no absolute right or wrong answer to these items, I am looking for the most commonly occurring pattern of responses for each one.

Please do not spend too much time thinking about each item – just go with your first instinct on the rankings. Please do not give a joint ranking for any of the items.

When you have finished, please return the consent form and question form to me via the box in the departmental office by Monday 12 March 2007. I will then issue your subject hour form and put that in your pigeonhole when I collect the forms.

Please note that this study is for ½ subject hour, so you should spend no more than ½ hour on this task.

Thanks for your help!

Leesa Clarke

[l.clarke@psych.york.ac.uk](mailto:l.clarke@psych.york.ac.uk)

1. Sue asked for an orange juice and lemonade. Jack ordered a diet coke with no ice.

Where were they?

Bar  
Restaurant  
Supermarket  
Zoo

2. At last it was time for dinner. Tigger was hungry and ate all of the food in his bowl. He curled up in his basket and licked his paws.

What was Tigger

Hamster  
Baby  
Cat  
Dog

3. Mum asked the woman if she could see what was wrong with Nathan's eye. It was very red, itchy and sore. She thought he might need some cream to make it better.

Where were they?

Hospital  
Hairdresser  
Doctors  
Vet

4. The cat did not like being put into the box and made lots of noise. Mum shut the box, and carried it to the car.

Where were they going?

Cattery  
Holiday  
Vet  
Swimming

5. Anna said she wanted it shorter for the summer. The lady picked up her scissors and began to cut.

Where was Anna?

Hospital  
Hairdresser  
Farm  
Tailor

6. Mum asked for the cod and chips. Dad wanted the haddock with mushy peas.

Where were they?

Fish and chip shop

Fishmonger

Supermarket

Dentist

7. Jack was at work. He opened the book at the right page and picked up the trumpet.

What was Jack's job?

Chef

Charity Worker

Student

Musician

8. They stood in a line, waiting for the gun to fire. Then it was time to run!

What were they doing?

Cooking

War Games

Athletics

Horse Racing

9. The music was so loud they couldn't hear themselves speak. They pushed their way through the crowd and tried to get to the front.

Where were they?

Pop Concert

Hospital

Nightclub

Hotel bar

10. Callum had never seen so many books. There were people working and no-one made a sound.

Where was Callum?

Factory

Office

Bookshop

Library

11. Mum pressed the “on” button and the music started playing. She turned up the volume and started to sing.

What did Mum switch on?

Television  
CD Player  
Radio  
Microwave

12. Jack felt terrible, and went to lie down under the duvet. His tummy hurt and all he wanted to do was sleep.

Where was Jack?

In the lounge  
In the garden  
In his bedroom  
At a friend’s house

13. The thunder was very loud and the lightning flashed all around. Sarah hid under her blanket and waited for the storm to pass.

How did Sarah feel?

Worried  
Scared  
Happy  
Cold

14. John rode as fast as he could, but he was not going to make it up the hill. In the end he stopped for a rest, then he pushed it the rest of the way to the top.

What was John riding?

Horse  
Tractor  
Bicycle  
Motorbike

15. The sun was very hot indeed. Holly had forgotten to bring her sun cream, and she was worried that she was going to burn. She covered herself up in her towel.

Where was Holly?

Beach  
Swimming pool  
Park  
Shops

16. Bob undid the screws and opened the back of the television. He could see straight away why they could not switch it on. He got his tools and started to fix the wires.

What did Bob do for a living?

Plumber  
Electrician  
Handyman  
TV Repair man

17. It was mum's 40<sup>th</sup> birthday and Dad had planned a surprise for her. He booked a big room at a hotel, and had invited all her friends.

What sort of surprise had dad planned?

Concert  
Cake  
Dinner  
Party

18. John was miserable. His throat hurt, his nose was running and he had a bad cough. Mum made him some honey and lemon to drink.

What was wrong with John?

Flu  
Hayfever  
Cold  
Tummy bug

19. Charlotte could hear something rustling in the bushes. Then she saw a stripey ginger tail poking out from underneath.

What was hiding in the bush?

Mouse  
Cat  
Dog  
Brother

20. Ali had been running for a long time. He was in front, but he was getting tired. He still had three miles to go before he could have a rest.

What was Ali taking part in?

A marathon  
Athletics training  
Cross country  
Netball practice

21. It was a hot summer's day and Hannah was weeding in her garden. She could hear something buzzing and it was getting louder. She was scared and looked around in a panic.

What did Hannah think the buzzing was?

Aeroplane  
Dog  
Wasp  
Fly

22. The doorbell rang, and a man in a red jacket waited at the door. The parcel was too big to fit through the letterbox, and he needed someone to sign for it.

Who was at the door?

Postman  
Courier  
Paper boy  
Milkman

23. They all stood around the table and sang while Connor blew out the candles. Then it was time to eat the cake his mum had made, and to start playing musical chairs.

Where were they all?

The zoo  
The pub  
At school  
Connor's house

24. The warm water splashed down on Holly's head. She got some shampoo and rubbed it into her hair.

Where was Holly?

In the shower  
In the bath  
At the cinema  
In the rain



25. John could see smoke coming from the house. He got his mobile phone and dialed '999'.

Who was John phoning?

Police  
Doctor  
His cousin  
Fire brigade

26. John fell over and tore a big hole in the knee of his new jeans. Mum got a needle and thread.

What was mum going to do?

Put stitches in John's knee  
Repair John's jeans  
Do some sewing  
Cook dinner

27. It was time for their summer holiday at last. The whole family was going to Disneyland in Florida, and they were on their way at last.

How were they travelling to Florida?

Tractor  
Plane  
Boat  
Car

28. It was late on Friday night, and Emma was having a party. Suddenly all the lights went out and the music stopped. Everyone looked around at each other in surprise.

What had happened?

Dinner was ready  
The neighbours complained  
It was bedtime  
A power cut

29. The nurse came in pushing a trolley. She told Mum that it was time for her to take some tablets.

Where were they?

Hospital  
Clinic  
Old people's home  
Airport

30. Abi was tired and hungry, and the sun was getting hot. They had been walking since breakfast and she had had enough. Dad said it was another 2 miles to the nearest village.

Where were they?

In London  
In the jungle  
In the mountains  
In the countryside

31. John was really pleased with himself – everyone kept telling him how good he looked and his trousers were getting very baggy.

Why was John pleased?

He wanted some new trousers  
He was on holiday  
He had lost weight  
He had been to the gym

32. They had been waiting for over an hour to see the elephants. They came quite close to the car and Freddie couldn't believe how big they were!

Where was Freddie?

Circus  
Safari Park  
Church  
Zoo

33. Adam put his foot down and turned the wheel. Everybody laughed as he crashed into Mum and Dad. Dad put his foot down and went after Adam to get his own back.

Where was Adam

At the funfair  
In a bumper car  
On the motorway  
In the supermarket

34. It was a hot summer's day and Ella was helping dad to sell some of her old toys. A boy came up to the table, and asked her if there was anything he could buy for a pound.

Where was Ella?

Dentist  
Front garden  
Church Hall  
Car boot sale

35. Bob was really annoyed. All of the leaves of his lettuces were gone and the cabbages were half eaten.

What had eaten them?

Fish  
Caterpillars  
Children  
Insects

36. Tom clicked the mouse and waited for the page to load. He hoped that it would help him with his history homework.

What was Tom using for his homework?

Computer game  
DVD  
Electronic book  
Internet site

37. It was very cold and there was frost on the ground. Jack wondered if there would be school tomorrow, because the weather forecast was for heavy showers.

What was expected to fall overnight?

Sleet  
Snow  
Hail  
Rain

38. John sealed the envelope and put a stamp on the front. He hoped that his friend would be glad to hear all his news.

What was John sending?

Email  
Letter  
Card  
CD

39. The referee blew his whistle. The ball was in the back of the net, and the home team was one up.

What were they playing?

Football  
Hockey  
Netball  
Darts

40. The car was making a terrible noise, and Ryan decided that it was not safe to carry on driving. He decided to pull over and ring for help.

Who was Ryan ringing?

His little sister  
A policeman  
A garage  
A breakdown company

41. After Emily's dad had finished his speech, Ben and Emily cut the cake while everyone took pictures. It had been a wonderful day!

What happened that day?

They won a cake  
They had a party  
They got married  
They went swimming

42. Clare held her nose and jumped in. The water was freezing and she let out a yell. Mum sat on the side with her feet in the cold water.

Where were they?

At the swimming pool  
In a forest  
In the shower  
At the seaside

43. It was their first night in the tent. The rain was pouring down and Ben's sleeping bag was getting wet. He wanted to go in the house but thought they would laugh at him.

Where was the tent?

In the woods  
On the beach  
In the garden  
At the camp site

44. Oliver had been feeling really sick, and he had a very high temperature. Mum decided to ring and see if someone could come and see him.

Who did mum ring?

Vet  
Doctor  
Bank  
Chemist

45. The man was carrying a bag of jewellery and lots of cash when the police stopped him. The watches and the rings had been stolen and did not belong to him.

Who was the man?

Mugger  
Burglar  
Jeweller  
Vicar

Thank you for your help – please return the completed sheets to the box in the departmental office.

## **Test Scenarios**

In Experiment 6, participants were administered test items divided into two testing sessions. The order of item set presentation was randomized. Two practice items were administered at the beginning of each test session, followed by the 16 test items administered in random order. The answers to the questions are listed in order of plausibility/coherence.

### **Item set 1**

1. (Practice Item) Sue asked for an orange juice and lemonade. Jack ordered a diet coke with no ice.

Where were they?

Bar  
Restaurant  
Supermarket  
Zoo

2. (Practice Item) Mum asked for the cod and chips. Dad wanted the haddock with mushy peas.

Where were they?

Fish and chip shop  
Fishmonger  
Supermarket  
Dentist

3. Clare held her nose and jumped in. The water was freezing! Mum sat on the side with her feet in the cold water.

Where were they?

At the swimming pool  
At the seaside  
In the shower  
In a forest

4. The warm water splashed down on Holly's head. She got some shampoo and rubbed it into her hair.

Where was Holly?

In the shower  
In the bath  
In the rain  
At the cinema

5. John sealed the envelope and put a stamp on the front. He hoped that his friend would be glad to hear all his news.

What was John sending?

Letter  
Card  
CD  
Email

6. It was mum's 40<sup>th</sup> birthday and Dad had planned a surprise for her. He booked a big room at a hotel, and had invited all her friends.

What was the surprise?

Party  
Dinner  
Cake  
Concert

7. The doorbell rang, and a man in a red jacket was at the door. The parcel was too big to fit through the letterbox.

Who was at the door?

Postman  
Delivery man  
Paper boy  
Milkman

8. Abi was tired and hungry, and the sun was getting hot. They had been walking since breakfast. Dad said it was another two miles to the nearest village.

Where were they?

In the country  
In the mountains  
In the jungle  
In London

9. They had been waiting for over an hour to see the elephants. They came quite close to the car and Freddie couldn't believe how big they were!

Where was Freddie?

Safari park  
Zoo  
Circus  
Church

10. At last it was time for dinner. Tigger was hungry and ate all of the food in his bowl. He curled up in his basket and licked his paws clean.

What was Tigger?

- A cat
- A dog
- A hamster
- A baby

11. Ali had been running for a long time. He was in front, but he was getting tired. He still had three miles to go, before he could rest.

What was Ali taking part in?

- A marathon
- Cross country
- Athletics
- Netball

12. Jack felt terrible, and went to lie down under his duvet. His tummy hurt and all he wanted to do was sleep.

Where was Jack?

- In his bedroom
- In the lounge
- At a friend's house
- In the garden

13. The music was so loud they couldn't hear themselves speak. They pushed their way through the crowd and tried to get to the front.

Where were they?

- Pop concert
- Nightclub
- Bar
- Hospital

14. The car was making a terrible noise. Ryan decided it was not safe to carry on driving. He pulled over and rang for help.

Who did Ryan ring?

- Breakdown company
- A garage
- A policeman
- His little sister



15. John fell over and tore a big hole in the knee of his new jeans. Mum got a needle and thread.

What was mum going to do?

Fix John's jeans  
Do some sewing  
Stitch John's cut knee  
Cook dinner

16. John rode as fast as he could, but could not make it up the steep hill. He stopped for a rest, then he pushed it the rest of the way to the top.

What was John riding?

Bicycle  
Motorbike  
Horse  
Tractor

17. Oliver had been feeling very sick, and he had a high temperature. Mum decided to ring and ask someone to come and see him.

Who did mum ring?

The doctor  
The chemist  
The vet  
The bank

18. Mum asked the woman to look at Nathan's eye. It was red, itchy and sore. She thought he might need some cream to make it better.

Where were they?

Doctors  
Hospital  
Vet  
Hairdressers

## **Item Set 2**

1. (Practice Item) Dad wanted to buy something better with a bigger engine. Mum said she wanted a bigger boot and a good radio.

What were they buying?

Car  
Van  
Caravan  
Dog

2. (Practice Item) Tom clicked the mouse and waited for the page to come up. He hoped it would help him with his history homework.

What was Tom using for his homework?

Internet  
Playstation  
DVD  
Magazine

3. It was their first night in the tent. The rain was pouring down and Ben's sleeping bag was wet. He wanted to go in the house, but thought they would laugh at him.

Where was the tent?

In the garden  
At the camp site  
In the woods  
On the beach

4. Bob was really cross. All of the leaves of his lettuces were gone and the cabbages were half eaten.

What had eaten them?

Caterpillars  
Insects  
Children  
Fish

5. They all stood around the table and sang as Connor blew out the candles. Then it was time to eat the cake his mum had made, and to start playing musical chairs.

Where were they all?

Connor's house  
The pub  
At school  
At the zoo

6. Callum had never seen so many books. There were people working and no-one made a sound.

Where was Callum?

Library  
Bookshop  
Office  
Factory

7. After Emily's dad had finished his speech, Ben and Emily cut the cake. Everyone took pictures. It had been a wonderful day!

What happened that day?

They got married  
They had a party  
They won a cake  
They went swimming

8. The referee blew his whistle. The ball was in the back of the net, and the home team was one goal up.

What were they playing?

Football  
Hockey  
Netball  
Darts

9. The man was carrying a bag of jewellery and cash when the police stopped him. The watches and the rings had been stolen and did not belong to him.

Who was the man?

A mugger  
A burglar  
A jeweller  
A vicar

10. John could see smoke coming from the house. He got his mobile phone and dialed '999'.

Who was John phoning?

Fire brigade  
Police  
Doctor  
Grandma

11. Anna said she wanted it shorter for the summer. The lady picked up her scissors and began to cut.

Where was Anna?

Hairdressers  
Tailor  
Hospital  
Farm

12. The thunder was very loud and the lightning flashed all around. Sarah hid under her blanket and waited for the storm to pass.

How did Sarah feel?

Scared  
Worried  
Cold  
Happy

13. John was really pleased with himself. Everyone kept telling him how good he looked and his trousers were getting very baggy.

Why was John pleased?

He had lost weight  
He had been to the gym  
He wanted new trousers  
He was on holiday

14. It was a hot summers' day and Hannah was weeding the garden. She could hear something buzzing and it was getting louder. She was scared and looked around to see what it was.

What did Hannah think the buzzing was?

Wasp  
Fly  
Plane  
Dog

15. It was time for their summer holiday at last! The whole family was going to Disneyland in Florida.

How were they travelling to Florida?

Plane  
Boat  
Car  
Tractor

16. The sun was very hot indeed. Holly had forgotten to bring her sun cream. She covered herself up in her towel.

Where was Holly?

Beach  
Pool  
Park  
Shops

17. The cat did not like being put into the box and made lots of noise. Mum shut the box, and carried it to the car.

Where were they going?

Vet  
Cattery  
Holiday  
Swimming

18. Charlotte could hear something moving in the bushes. Then she saw a stripey ginger tail poking out.

What was hiding in the bush?

Cat

Dog

Mouse

Brother

Hatcher Grading and *WORD* Reading Age Equivalents for Experiment 6 Scenarios.

Scenario Name	Hatcher Grade	<i>WORD</i> reading age equivalent (years)
Nathan's eye	5	6.1
Concert	6	6.1
Tummy ache	3	<6
Bike ride	6	6.1
Mum's birthday	5	6.1
Marathon run	5	6.1
Postman	3	<6
Shower	2	<6
Hole in jeans	3	<6
Long walk	5	6.1
Safari park	6	6.1
Sending a letter	4	6
Car breakdown	4	6
Jumping in the pool	3	<6
Calling the doctor	4	6
Hungry cat	4	6
Cat in a box	4	6
Library	4	6
Thunderstorm	4	6
No sun cream	3	<6
Cat in the bushes	3	<6
Buzzing bee	7	6.2
Birthday cake	7	6.2
Fire!	2	<6
Disneyland trip	3	<6
Baggy trousers	5	6.1
Chewed lettuces	4	6
Goal!	3	<6
Wedding	5	6.1
Camping in the garden	7	6.2
Burglar	6	6.1
Haircut	2	<6
Mean Book Grade for scenarios (SD)	4 (1.44)	6